

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

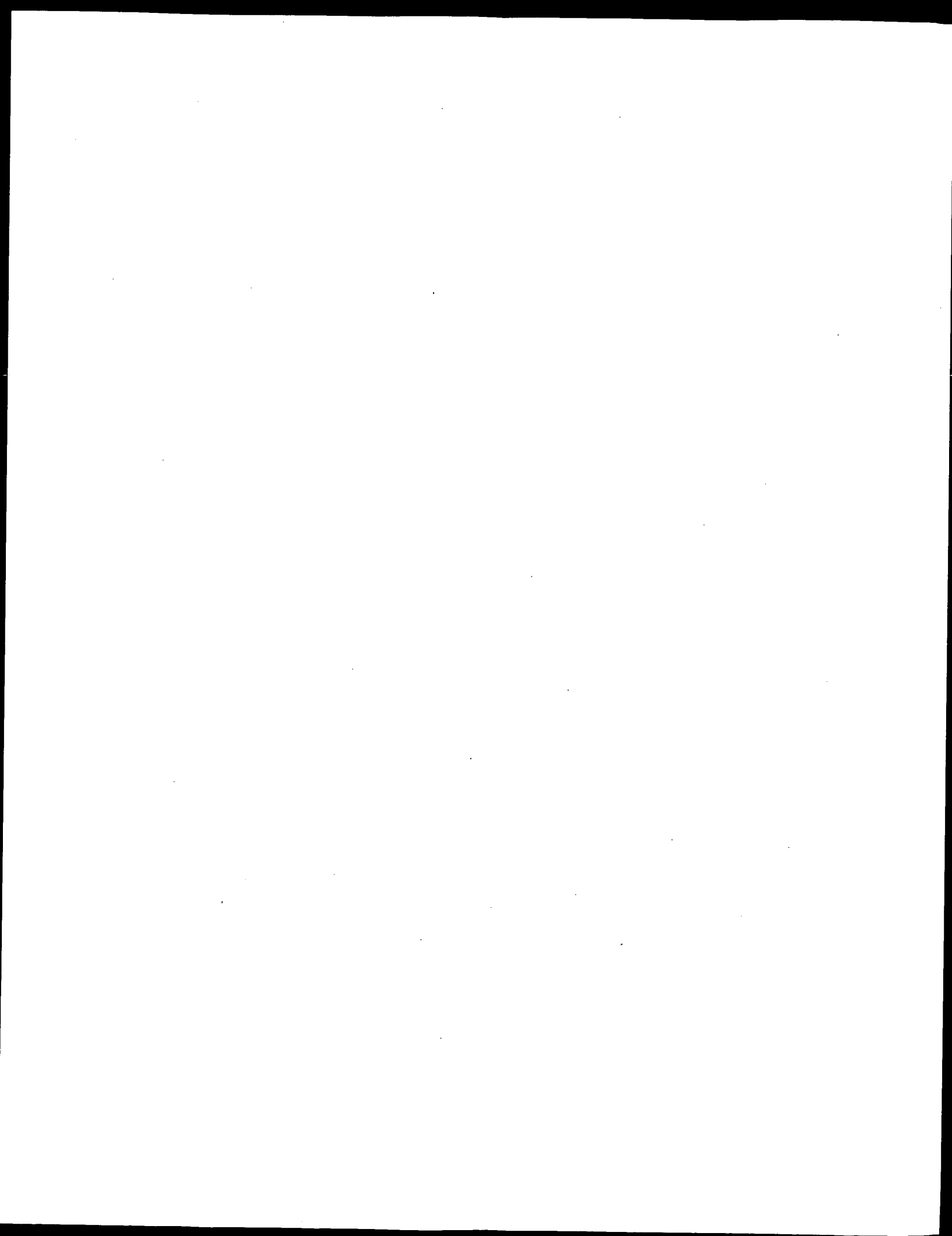
Searcher: P. Schreiber
 Searcher Phone #: 308-4292
 Searcher Location: CM112E18
 Date Searcher Requested: _____
 Date Completed: 11/14
 Searcher Prep & Review Time: 6
 Client Prep Time: _____
 Filing Time: 10

Type of Search

NA Sequence (#) 3
 AA Sequence (#) 1
 Structure (#) _____
 Bibliographic _____
 Litigation _____
 Fulltext _____
 Patent Family _____
 Other _____

Vendors and cost where applicable

STN _____
 Dialog _____
 Questel Orbit _____
 CompuLink _____
 Lexis Nexis _____
 Sequence Systems CompuLink
 WWW Internet _____
 Other (specify) _____



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2001, 09:53:10 ; Search time 1462.83 Seconds
(without alignments)
12355.404 Million cell updates/sec

Title: US-09-589-510-3

Perfect score: 1912
Sequence: 1 acccagcgctccgcaattt.....aaaaaaaaaaaaaaaa 1912

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb.est1:*
2: gb.est2:*
3: gb.est3:*
4: gb.est4:*
5: gb.est5:*
6: gb.est6:*
7: gb.est7:*
8: gb.est8:*
9: gb.est9:*
10: gb.est10:*
11: gb.est11:*
12: gb.est12:*
13: gb.est13:*
14: gb.est14:*
15: gb.est15:*
16: gb.est16:*
17: gb.est17:*
18: gb.est18:*
19: gb.est19:*
20: gb.est20:*
21: gb.est21:*
22: gb.est22:*
23: gb.est23:*
24: gb.est24:*
25: gb.est25:*
26: gb.est26:*
27: gb.est27:*
28: gb.est28:*
29: gb.est29:*
30: gb.est30:*
31: gb.est31:*
32: gb.est32:*
33: gb.est33:*
34: gb.est34:*
35: gb.est35:*
36: gb.est36:*
37: gb.est37:*
38: gb.est38:*
39: gb.est39:*
40: gb.est40:*
41: gb.est41:*
42: gb.est42:*
43: gb.est43:*

44: em.esthum10:*
45: em.esthum11:*
46: em.esthum12:*
47: em.esthum13:*
48: em.esthum14:*
49: em.esthum15:*
50: em.esthum16:*
51: em.esthum17:*
52: em.esthum18:*
53: em.esthum19:*
54: em.esthum20:*
55: em.esthum21:*
56: em.esthum22:*
57: em.esthum23:*
58: em.esthum24:*
59: em.esthum25:*
60: em.esthum26:*
61: em.esthum27:*
62: em.esthum28:*
63: em.esthum29:*
64: em.esthum30:*
65: em.esthum31:*
66: em.esthum32:*
67: em.esthum33:*
68: em.esthum34:*
69: em.esthum35:*
70: em.esthum36:*
71: em.esthum37:*
72: em.esthum38:*
73: em.esthum39:*
74: em.esthum40:*
75: em.esthum41:*
76: em.esthum42:*
77: em.esthum43:*
78: em.esthum44:*
79: em.esthum45:*
80: em.esthum46:*
81: em.esthum47:*
82: em.esthum48:*
83: em.esthum49:*
84: em.esthum50:*
85: em.esthum51:*
86: em.esthum52:*
87: em.esthum53:*
88: em.esthum54:*
89: em.esthum55:*
90: em.esthum56:*
91: em.esthum57:*
92: em.esthum58:*
93: em.esthum59:*
94: em.esthum60:*
95: em.esthum61:*
96: em.esthum62:*
97: em.esthum63:*
98: em.esthum64:*
99: em.esthum65:*
100: em.esthum66:*
101: em.esthum67:*
102: em.esthum68:*
103: em.esthum69:*
104: em.esthum70:*
105: em.esthum71:*
106: em.esthum72:*
107: em.esthum73:*
108: em.esthum74:*
109: em.esthum75:*
110: em.esthum76:*
111: em.esthum77:*
112: em.esthum78:*
113: em.esthum79:*
114: em.esthum80:*
115: em.esthum81:*
116: em.esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	% Query Match	Length	ID	Description
1	597	31.2	1595	192	AK012170 MMS MUSCU
2	591.6	30.9	610	136	BE512278 94606BD9
3	524	27.4	1427	104	AJ131743 B927 097
4	515	26.9	555	122	AW927556 945018G05
5	494.6	25.9	717	154	BC464635 EMI_72_BO
6	476.8	24.9	512	122	AM927557 945018G05
7	469	24.5	549	136	BE512425 946071A06
8	466	24.4	474	136	BE512426 946071A06
9	463.8	24.3	994	162	BE035216 MO01E02 M
10	457.8	23.9	994	162	BE035216 MO01E02 M
11	456.8	23.9	535	154	BE513383 AL553383
12	448.6	23.5	764	174	BC464307 EMI_72_BO
13	448	23.4	919	106	BE126632 EST472278
14	433.2	22.7	962	106	AL542244 AL542244
15	427	22.3	893	23	AL557272 AL557272
16	424.8	22.2	862	106	A1667714 605026D04
17	422.8	22.1	877	106	AL559327 AL559327
18	419.2	21.9	632	104	AL535830 AL535830
19	419	21.9	899	105	BE204244 EST36920
20	416	21.8	626	113	AL514734 AL514734
21	403.6	21.1	1066	162	BE035218 BE035218
22	400.4	20.9	664	24	BE176943 BNLGH1689
23	399	20.9	564	149	BF473384 MO01E04 M
24	397.2	20.8	695	24	BF473384 MO0E03_F
25	382.8	20.0	939	141	BF473384 BNLGH1131
26	378.6	19.8	866	106	A1725799 BNLGH1131
27	378.4	19.7	876	24	BE501271 601675912
28	376.2	19.7	899	153	AL551666 AL551666
29	374.6	19.6	476	164	A1722028 F6118804_Y
30	367.4	19.2	1012	139	BG397336 602439206
31	365	19.1	576	156	BE492538 WHE0534_A
32	357.8	18.7	440	136	BE492538 WHE0534_A
33	357.8	18.7	653	20	BE745673 BE745673
34	357.6	18.7	653	20	BE495698 EST415242
35	355	18.6	646	20	BE495698 WHE1279_H
36	353.8	18.5	503	117	BF957572 602246315
37	353.6	18.5	503	117	A1444686 486015B10
38	345.6	18.1	757	106	A1444686 486015F09
39	343.6	17.9	728	139	AL560969 AL560969
40	341.2	17.9	728	139	AW566092 660061G12
41	340	17.8	782	154	AW566092 660061G12
42	339.4	17.8	400	173	BE745221 601573974
43	339	17.7	1005	139	BE745221 601573974
44	338.8	17.7	698	116	BG408569 602529592
45	333.8	17.5	606	103	BE120437 AU120437
					BG101744 EMI_5_G05
					BE733442 601568827
					AM460585 da26d01.x
					A1878120 fc57502.y

RESULT	1
LOCUS	AK012170
DEFINITION	AK012170 1595 bp mRNA HTC 08-FEB-2001 Mus musculus 10 days embryo cDNA, RIKEN full-length enriched library, clone:2610528G06, full insert sequence.
ACCESSION	AK012170
VERSION	AK012170.1 GI:12848750
KEYWORDS	CAP trapper.
SOURCE	Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA, clone_1b5:RIKEN full-length enriched mouse cDNA library clone:2610528G06.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathli; Muridae; Murinae; Mus.
AUTHORS	1 (sites)
TITLE	Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning

[illegible]

	Query Match	26.9%	Score 515;	DB 122;	Length 555;
	Best Local Similarity	95.5%;	Pred. No. 1.4e-117;		
	Matches 530;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0;
213	gggcacagcgcgcgcgcgacgagcgccgctggcggcgcgcacatgatcgcagaagaa 272				

```

Seq primer: POLYTMix
High quality sequence start: 23
High quality sequence stop: 680
POLYA-No.
Location/Qualifiers
1..717
FEATURES
source
```

/organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Embryo 1 (EM1)"
 /note="Organ: Embryos germinated for 24 hr. Vector:
 pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
 EcoRI. The library was made from poly(A RNA in the cloning
 vector lambda Zap II. Clones to be sequenced were
 prepared by mass excision."

BASE COUNT 197 a 127 c 182 g 210 t 1 others

ORIGIN

Query Match 25.9%; Score 494.6; DB 154; Length 717;
 Best Local Similarity 84.2%; Pred. No. 1.8e-112;
 Matches 630; Conservative 1; Mismatches 76; Indels 41; Gaps 5;

QY 955 gatgaagaattgcagagctgtacactggtgtttgttcaatgataaggtccacatgtg 1014
 |||||||
 DB 11 GAGGAGGAATTCGAGAGCTGTACCTGCTGCTTTGTCATTCATGATGAGGTCACATGTTG 70
 |||||||

QY 1015 gatacgaatgtttttcttcttaacacgtgcatgagagcccatcaccacatgtg 1074
 |||||||
 DB 71 GATATCGATGCTTTCTTCTTATCTTAATCGTCACTGAGAGCCCTTTATCCACATTTG 130
 |||||||

QY 1075 atactgttacaataagggagataatgtaatgtaaggaaggaatgataatgcaagtccacat 1134
 |||||||
 DB 131 ATACTTGTACGATAGGAGGAAATGTAATGTAAGAGCAATGATATGACAAATCCACAT 190
 |||||||

QY 1135 ggtacacgtgtgactcttctagataagctgtgtatcttctgagcagagacataatggccct 1194
 |||||||
 DB 191 GGTATACGAGTGTGATCTTCTTGTATGATAGGTTGCTGATTTTGTGAGCAACATATGCCCCCT 250
 |||||||

QY 1195 actgagatgatacagatatgtgctatcccgagcacaagtggagagatgatalgtatgaa 1254
 |||||||
 DB 251 GCTGAGATGATACGATTTTGGCAATCGACACACAGTGGAGGAGATTGATATCGATGAA 310
 |||||||

QY 1255 gaaagtctgtcttattgagcgagatcgagacagacacatcttgaacatgcatatcaa 1314
 |||||||
 DB 311 GAAAGCTTGTATTTAGGCGAGATCGACACAGACATCTTTGAGACATGATTTTACG 370
 |||||||

QY 1315 ttgataaccctgcagaggtgtgtctcaagaactaaggaagaagaatctgcaagct 1374
 |||||||
 DB 371 TTGCTATACCTGCGCAGCTGTGTCANAGACTAATGSAAGAAAGATGTCGAGGCT 430
 |||||||

QY 1375 gatctcgaagaatcagtggtgtctatcttggatgacaaatccctggctggtctccag 1434
 |||||||
 DB 431 GATCTTGAGGAAGTTAGTGTCTCTATTGGAATGCAATCTTGGCTGCTCTCCAG 490
 |||||||

QY 1435 gagcaacaagaatatacatcaactagatttgatctctctgtcgtggaagctcgaagag 1494
 |||||||
 DB 491 GACCAACAGACAGATACATCACTAGATT---TCTCTGTTGTAGAAAGCTCGAAGAA 546
 |||||||

QY 1495 aatgtatgttcagacgtcgaagaatcatcatgtagtcatgctctgctcacaagttcaatgctc 1554
 |||||||
 DB 547 AATGTATGTGCCAGCTC---AAGGCCATGTACATCGATCGATCTCATATCTGGGGAG 603
 |||||||

QY 1555 tactgtgtcttgagcgagacacatttcggggggaagcgctgaatttgcagtgctgctgt 1614
 |||||||
 DB 604 TTTATTTTGTAGTGCACGCTTAT-----GGCTT 631
 |||||||

QY 1615 gtgttagtcccaagaagaatgtgttcgcgcatatgtcgttcaacgactgttcgtg 1674
 |||||||
 DB 632 CGGTACTGTCCAGAGAGGACTGTGTCGGTATATGCTGTCCACATTA-TGTTCCGCTN 690
 |||||||

QY 1675 atgagatgtcacccggtgcaggaatg 1702
 |||||||
 DB 691 A-TGATTGACAATGTGTGCAGGAATTG 717
 |||||||

RESULT 6
 AM927557 512 bp mRNA EST 30-MAY-2000
 LOCUS
 DEFINITION 945018G05.X2 945 - Mixed adult tissues from Walbot lab, same as 707

ACCESSION (SK) Zea mays cDNA, mRNA sequence.
 AM927557
 VERSION AM927557.1 GI:8102879
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 512)

REFERENCE
 AUTHORS Walbot, V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 JOURNAL University
 COMMENT Unpublished (1999)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 945018 row: G column: 05.
 Location/Qualifiers
 1. 512
 /organism="Zea mays"
 /cultivar="W23"
 /db_xref="taxon:4577"
 /clone_lib="945 - Mixed adult tissues from Walbot lab,
 same as 707 (SK)"
 /tissue_type="tassel, kernel, silk, husk, root, leaf"
 /dev_stage="fully-grown"
 /lab_host="DH10b"
 /note="Organ: tassel, kernel, silk, husk, root, leaf;
 Vector: pCDAD10; Site 1: EcoRI; cDNA library from fully
 differentiated maize tissues from an active Mutator plant.
 Tissue ratio is 4:2:1:1:1 (tassel, kernel, silk, husk,
 root, leaf). Unidirectionally cloned. New library number
 given to library 707 for additional sequencing."

BASE COUNT 148 a 98 c 151 g 115 t

ORIGIN

Query Match 24.9%; Score 476.8; DB 122; Length 512;
 Best Local Similarity 95.7%; Pred. No. 4.6e-108;
 Matches 490; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 230 gggagggcgccgggctggtgcgacatgctgcgcagaagaagatgcccggcgcg 289
 |||||||

DB 1 GCGAGGCGCCGGGCTGGCGGTGACATGATCCGCCAGAGAAGATGCCGTCGCGCGC 60
 |||||||

QY 290 tgcctcttggtgtccgcgcgcagcggaagagcggtgagcgtggtggaatgcccag 349
 |||||||

DB 61 TGCTCTTGGCGGCGCGCCGCGGCGGCGCAAAACGGCTGACGCTGCGCATAGCGCAGG 120
 |||||||

QY 350 agctggcgcaaggtcccttctctcctatgtagtgagatcgaagatgtactcctcgagg 409
 |||||||

DB 121 AGCTCGGACCAAGTCCCATTTCTGTCTATGTGATGATGCAAAATGTATACCTTATTAAGG 180
 |||||||

QY 410 tcaagaanaatgaggtgctgtaglaagaaatttcogtagagctataagtttggtataaagg 469
 |||||||

DB 181 TCAAGAAATGTAGGTGCTGATGAGAAATTTCCGTAGACCTTATGCTTATTAAGG 240
 |||||||

QY 470 aaaaacaagaagttatgaaagagaggttactggaacttcccaagaagagctggaagta 529
 |||||||

DB 241 AAAACAAGAGGTTTATGAAGGAGAGTTTGAACCTTTCCCCAGAGAGGCTGAGACA 300
 |||||||

QY 530 caactgttgatatagcaaaaagcatatgacatgtaactcaagctaaagactgttaaag 589
 |||||||

DB 301 CAATGTGATATGCAAAAGCATATAGCCACTATTCATTTGCTTAAGACTGTCAAG 360
 |||||||

QY 590 ggaactaagaactgaagtagatcttcaattatgtagtctgtagtcaaggaagaagtg 649
 |||||||

DB 361 GGACTAAGCAATTTGAATGACGCTTCAATTTATGATGCTTTGATCAAGGAAAAGGTGG 420
 |||||||

QY 650 cagtggtgtagtatacatcgaagcaaatgagtgagcaatgagaagatgtatag 709
 DB 421 CAGTGGGTGATGTTATACATGGAAGCAAAATGAGACAGTGAAGAGTTGGTAGAT 480
 QY 710 gtagctctttgtacagaatcagctctgaa 741
 DB 481 GTGATCTTTTGTCTACAGATATGATCTTGAA 512

RESULT 7
 BE512425/c
 LOCUS BE512425 549 bp mRNA EST 07-AUG-2000
 DEFINITION 946071A06.x1 946 - tassel primordium prepared by schmidt lab Zea
 mays cDNA, mRNA sequence.
 ACCESSION BE512425
 VERSION BE512425.1 GI:9733673
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 Clade; Panicoidae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 549)
 AUTHORS Walbot, V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 946071 row: A column: 06.
 Location/Qualifiers
 1. 549
 /organism="Zea mays"
 /cultivar="OH43"
 /db_xref="taxon:4577"
 /clone_lib="946 - tassel primordium prepared by schmidt
 lab"
 /issue_type="tassels"
 /dev_stage="just after the transition from vegetative to
 inflorescence development"
 /lab_host="XLOLR"
 /note="Organ: tassels; Vector: HybridZAP; Site: 1: EcoRI;
 Site: 2: XhoI; George Chuck dissected immature tassels
 between 1mm and 3mm. Sharon Stanfield prepared the cDNA
 library in HybridZAP. Sample insert size range was 350 bp
 to 3 Kb with a 1 kb average."
 BASE COUNT 150 a 146 c 116 g 137 t
 ORIGIN

Query Match 24.5%; Score 469; DB 136; Length 549;
 Best Local Similarity 94.5%; Pred. No. 4; le-106;
 Matches 534; Conservative 1; Mismatches 11; Indels 19; Gaps 4;

QY 1210 atattggcattccgagcaagtgagagatgtatatgtatgaagaatctgtctat 1269
 DB 549 ATATTGGCATTCGAGCCAAAGTGGAGACATTTATGATGTAAGAAAGCTTGTAT 490
 QY 1270 ttaagcgagatcgacacagacatcttggagacatcttaaatgtatcaacctgc 1329
 DB 489 TTAGCGGAGATCGACACACACATCTTTAAGACATCTTACATTTGATATCACTGCC 430
 QY 1330 agcgtggtcctaagaactgaatgagaagaatctgcaaggctgattctcgagaagtc 1389
 DB 429 AGCGTGGTCTTAAAGACTAATGAGAGAAATATGCAAGCGTGTCTCGAGGAAGTC 370
 QY 1390 agtgggtctatttgatgccaatctcgtcgtcgtctgccagagacaagaaga 1449

DB 369 AGTGGGCTCATTTTGGATGCCAAATCCTCGGCTCGGCTGCCAGACCAAGAAAAGA 310
 QY 1450 tacatcaactagatttgatctctcgtctgtggaagtcctgaagaatgtatgtccagc 1509
 DB 309 TACATCACTTAATTTTGGATCACTGTCTGTGGAATCTCGAAGAAATGTAGTTGCAGC 250
 QY 1510 tcgaagatcatctagtcgtatctatctctcagagttcatagctactcgtcttgagc 1569
 DB 249 TCGAAAGTCACTAGTGAATGATGATCTGCTACA-----GCTCTTGGAGC 205
 QY 1570 gaggcatttcggggggaagcgctgaaatcttcagtgctcgtctgtatgctccaca 1629
 DB 204 GAGCAATTTTGGGGGGGAGACGCTTGAATTTTGCAGTCCGCTTGTAGTCTCCAA 145
 QY 1630 gaagacttgctccgacatctgctctgaagcagctgttcgtatgattgtgtccac 1689
 DB 144 -AAACTTGTGACCGCGCATTTGCTGTTCACGCACTGTTCGCTGATTAATTTGTCACC 86
 QY 1690 ggtgaggaatctccgtgtgtg-cttattctgtcatcagtgctcggagtc--tgcct 1746
 DB 85 GGTGAGGAATTTGCCGTGTGTTTATCTTCTCATCGTGTCCGGAATCTGTGCT 26
 QY 1747 ccacgggtgtgattggcccggaacc 1771
 DB 25 CCACGGGTGTATTGTGCCCGAACC 1

RESULT 8
 BE512426
 LOCUS BE512426 474 bp mRNA EST 07-AUG-2000
 DEFINITION 946071A06.y1 946 - tassel primordium prepared by schmidt lab Zea
 mays cDNA, mRNA sequence.
 ACCESSION BE512426
 VERSION BE512426.1 GI:9733674
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 Clade; Panicoidae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 474)
 AUTHORS Walbot, V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 946071 row: A column: 06.
 Location/Qualifiers
 1. 474
 /organism="Zea mays"
 /cultivar="OH43"
 /db_xref="taxon:4577"
 /clone_lib="946 - tassel primordium prepared by schmidt
 lab"
 /issue_type="tassels"
 /dev_stage="just after the transition from vegetative to
 inflorescence development"
 /lab_host="XLOLR"
 /note="Organ: tassels; Vector: HybridZAP; Site: 1: EcoRI;
 Site: 2: XhoI; George Chuck dissected immature tassels
 between 1mm and 3mm. Sharon Stanfield prepared the cDNA
 library in HybridZAP. Sample insert size range was 350 bp
 to 3 Kb with a 1 kb average."
 BASE COUNT 110 a 121 c 162 g 81 t
 ORIGIN

Wed Nov 14 08:34:11 2001

us-09-589-510-3.std.rst

Page 14

```

Oy 1571 agacacatttcggggg-----aaagcgcttgaatttgcagtgccctg-ctgtgttagt 1622
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69 AGCACGTTTCGCGCGGGAAGCAAAACTGTTGAAATTTGTATGACTCTTTTGTGAGC 10
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1623 ctccaraga 1631
      |||||: ||
Db 9 CTCACGCCA 1
      |||||

```

Search completed: November 13, 2001, 11:07:51
Job time: 4481 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 06:55:54 ; Search time 18.34 Seconds
(without alignments)
558.288 Million cell updates/sec

Title: US-09-589-510-4
Perfect score: 2263
Sequence: 1 MREEVOSTSKKRIATHTH.....YLDKSSARLLQEQERYIT 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1721.5	76.1	456	2	US-08-933-750C-5
2	1721.5	76.1	456	4	US-09-234-613-5
3	117	5.2	806	4	US-08-999-774A-6
4	111.5	4.9	327	4	US-09-446-504-3
5	111.5	4.9	652	4	US-08-987-123-2
6	108.5	4.8	686	2	US-08-993-228-12
7	104.5	4.6	723	4	US-08-928-213B-11
8	104.5	4.6	942	3	US-09-074-579-1
9	104.5	4.6	942	4	US-09-388-774-1
10	101.5	4.5	273	4	US-08-928-213B-10
11	100.5	4.4	1184	4	US-09-541-782-2
12	98	4.3	2101	1	US-08-466-390-4
13	98	4.3	2101	1	US-08-470-950-4
14	98	4.3	2101	1	US-08-467-781-4
15	98	4.3	2101	2	US-08-483-924-4
16	98	4.3	2101	4	US-09-452-294-1
17	97.5	4.3	456	4	US-09-268-364-21
18	97	4.3	2101	1	US-08-195-487-4
19	97	4.3	2101	1	PCT-US93-06160-4
20	96.5	4.3	876	1	US-08-717-515-4
21	96.5	4.3	1276	1	US-08-717-515-8
22	96	4.2	544	3	US-08-559-397A-29
23	95.5	4.2	1018	1	US-08-072-610-2
24	95.5	4.2	1018	2	US-08-719-822B-2
25	95.5	4.2	1018	4	US-09-092-458-2
26	94.5	4.2	1119	4	US-09-396-651B-2
27	93.5	4.1	479	4	US-09-446-504-64

28	93.5	4.1	545	2	US-08-467-822-30	Sequence 30, Appl
29	93.5	4.1	545	4	US-08-433-697-30	Sequence 30, Appl
30	93.5	4.1	545	4	US-08-466-248-30	Sequence 2, Appl
31	93	4.1	405	2	US-08-222-719-2	Sequence 2, Appl
32	93	4.1	405	2	US-08-470-925-2	Sequence 2, Appl
33	93	4.1	405	2	US-08-471-613-2	Sequence 2, Appl
34	93	4.1	405	5	PCT-US93-10443-2	Sequence 1, Appl
35	92.5	4.1	406	2	US-08-222-719-1	Sequence 1, Appl
36	92.5	4.1	406	2	US-08-470-925-1	Sequence 1, Appl
37	92.5	4.1	406	2	US-08-471-613-1	Sequence 1, Appl
38	92.5	4.1	406	5	PCT-US93-10443-1	Sequence 3, Appl
39	92.5	4.1	1964	2	US-08-790-912-3	Sequence 2, Appl
40	92.5	4.1	2052	2	US-08-790-912-3	Sequence 2, Appl
41	90.5	4.0	249	2	US-08-991-946A-1	Sequence 1, Appl
42	90.5	4.0	571	4	US-08-961-083-4	Sequence 4, Appl
43	90.5	4.0	834	1	US-08-471-033-21	Sequence 21, Appl
44	90.5	4.0	834	2	US-08-471-044-21	Sequence 21, Appl
45	90.5	4.0	834	2	US-08-463-483A-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-933-750C-5
Sequence 5, Application US/08933750C
Patent No. 5932442
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guebler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HMCINOT01

CLONE: 9476
US-08-933-750C-5

Query Match 76.1%; Score 1721.5; DB 2; Length 456;
Best Local Similarity 72.5%; Pred. No. 1.3e-154;
Matches 330; Conservative 72; Mismatches 52; Indels 1; Gaps 1;

```
QY 1 MKIEEVOSTSKKRIATHTHIKGLDANGMAIALAGFVGAAAREAGLAVDMIRKK 60
DB 1 MKIEEVKSTTKTORIASHSHVKGGLDESLAKQAASLVGQENARACGYIVELIESKK 60
QY 61 MAGRAVLLAGPAPGKTALALAGIAGELGSKVPCPCPVGSEVYSSEVKKTEVLMENFRRAI 120
DB 61 MAGRAVLLAGPAPGKTALALAGIAGELGSKVPCPCPVGSEVYSSEVKKTEVLMENFRRAI 120
QY 121 GLRIKENKEVEGEVTELSPEEASTTGAGYAKSISHVITSKTKVGTOKLIDSSYDAL 180
DB 121 GLRIKETKEVEGEVTELTPECTENPMGCGIKTISHVITGLKTKAGTKOLKIDPSIFESL 180
QY 181 IREKAVADVITYIANSAGVAVRGCDSPATEVDLEAEYVPIPKGEVHKKEIVODVTL 240
DB 181 QKEVERAGDVITYIANSAGVAVRGCDTPATEVDLEAEYVPLPKGDVHKKEIIVODVTL 240
QY 241 HDLDAANAPOGGODILSLMGOMMKPRKTEITEKLRQEIHKVNNRIYDGLAEVPGVLF 300
DB 241 HDLDAANARPOGGODILSMGOLMKPKKTEITDKLRGEINVKYIDOGIAELVPGVLF 300
QY 301 IDEVHMLDIECFSYLNRALESPLSPYIATNRGICNVRGTDMTSPHGIPLDLDRLVI 359
DB 301 VDEVHMLDIECFYTLHRALESIAPIVIFASNRCNVIKTEIDITSPHGIPLDLDRLVI 360
QY 360 IRTETGPTEMIOIATRAOYEIDMDESLAYIGEIGQTSLRHAIOIISPAVSVSKTN 419
DB 361 IRTMLTYQEMKQIKIRAOETGINISEALNHIGEIGTKTLRYSVOLTPANLLAKIN 420
QY 420 GREKICKADLEEVSGLYIDAKSSARLLQEOOERYI 454
DB 421 GKDSIEKEHVEISELFPYDAKSSAKILLADQODKYM 455
```

RESULT 2
US-09-234-613-5
Sequence 5, Application US/09234613
Patent No. 6132973
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/933,750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PP-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HMCINOT01
CLONE: 9476
US-09-234-613-5

Query Match 76.1%; Score 1721.5; DB 4; Length 456;
Best Local Similarity 72.5%; Pred. No. 1.3e-154;
Matches 330; Conservative 72; Mismatches 52; Indels 1; Gaps 1;

```
QY 1 MKIEEVOSTSKKRIATHTHIKGLDANGMAIALAGFVGAAAREAGLAVDMIRKK 60
DB 1 MKIEEVKSTTKTORIASHSHVKGGLDESLAKQAASLVGQENARACGYIVELIESKK 60
QY 61 MAGRAVLLAGPAPGKTALALAGIAGELGSKVPCPCPVGSEVYSSEVKKTEVLMENFRRAI 120
DB 61 MAGRAVLLAGPAPGKTALALAGIAGELGSKVPCPCPVGSEVYSSEVKKTEVLMENFRRAI 120
QY 121 GLRIKENKEVEGEVTELSPEEASTTGAGYAKSISHVITSKTKVGTOKLIDSSYDAL 180
DB 121 GLRIKETKEVEGEVTELTPECTENPMGCGIKTISHVITGLKTKAGTKOLKIDPSIFESL 180
QY 181 IREKAVADVITYIANSAGVAVRGCDSPATEVDLEAEYVPIPKGEVHKKEIVODVTL 240
DB 181 QKEVERAGDVITYIANSAGVAVRGCDTPATEVDLEAEYVPLPKGDVHKKEIIVODVTL 240
QY 241 HDLDAANAPOGGODILSLMGOMMKPRKTEITEKLRQEIHKVNNRIYDGLAEVPGVLF 300
DB 241 HDLDAANARPOGGODILSMGOLMKPKKTEITDKLRGEINVKYIDOGIAELVPGVLF 300
QY 301 IDEVHMLDIECFSYLNRALESPLSPYIATNRGICNVRGTDMTSPHGIPLDLDRLVI 359
DB 301 VDEVHMLDIECFYTLHRALESIAPIVIFASNRCNVIKTEIDITSPHGIPLDLDRLVI 360
QY 360 IRTETGPTEMIOIATRAOYEIDMDESLAYIGEIGQTSLRHAIOIISPAVSVSKTN 419
DB 361 IRTMLTYQEMKQIKIRAOETGINISEALNHIGEIGTKTLRYSVOLTPANLLAKIN 420
QY 420 GREKICKADLEEVSGLYIDAKSSARLLQEOOERYI 454
DB 421 GKDSIEKEHVEISELFPYDAKSSAKILLADQODKYM 455
```

RESULT 3
US-08-999-774A-6
Sequence 6, Application US/08999774A
Patent No. 6274312
GENERAL INFORMATION:
APPLICANT: Gish, Kurt C.
APPLICANT: Seghezzi, Wolfgang
APPLICANT: Shanahan, Frances
APPLICANT: Lees, Emma M.
TITLE OF INVENTION: Intracellular Regulatory Molecules;
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,774A
; FILING DATE: 10-DEC-1997
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,818
; FILING DATE: 11-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P. 34,090
; REGISTRATION NUMBER: US 60646
; REFERENCE/DOCKET NUMBER: DX0646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1200
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 806 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-999-774A-6

```

```

Query Match          5.2%; Score 117; DB 4; Length 806;
Best Local Similarity 25.1%; Pred. No. 0.021;
Matches 55; Conservative 36; Mismatches 68; Indels 60; Gaps 13;

```

```

QY 64 RAVLAGPPATGKTALALGIAOELSKVFCPMVGSSEYSEVKKTEVLMENFRAIGLR 123
   |::||| ||| ||| |::| |::| |::| |::| |::| |::| |::| |::|
DB 239 RGLILGPETGKTILARAVANETGAF--FLLNGPELMSKLAGSE--SNLRKAF--- 290
QY 124 IKENKE-----VYGEVTELSPEAESESTGGYAKSISHVITSLKTVGTQKLDSIY 177
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 291 --EEAEKNAPALIFIDELDAIPKR-EKTHGEVERIVSQLTL--MGLKOR----- 338
QY 178 DALIKEVANGDYIYEANS-----GAVKRVGRCDSPATEYDLBAEEVVP1PKG----- 226
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 339 -----AHYIVMAATRPNSIDPALRRFR-----FDREVDIGIPATGRLEIL 381
QY 227 EVH-KKKEIVQDYLHLDANAAQPOG--GODILSLMGQ 262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 382 QIHKKMKLADV---DLQEVANETHGHGADLALCSE 417

```

```

RESULT 4
US-09-446-504-3
; Sequence 3, Application US/09446504
; Patent No. 6218150
;
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshimi
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazuo
; APPLICANT: MUKAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/446,504
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24

```

```

; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
;
; US-09-446-504-3

```

```

Query Match          4.9%; Score 111.5; DB 4; Length 327;
Best Local Similarity 21.5%; Pred. No. 0.017;
Matches 79; Conservative 63; Mismatches 112; Indels 113; Gaps 18;

```

```

QY 54 DMIRCKMAGRA-----VLAGPPATGKTALALGIAOELSKVFCPMVGS----- 99
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 26 DIVGQEHIVKRLKHVKTGSMPLHLPAGPVGKTTAALAREL-----FGENMR 76
QY 100 ---EVYSSEVKKTEVLM---NFRRAIGLRIRKENKVEYGEVTELSPEAES---TTG 149
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 77 HNFLNMSDGRGINVIREKVEFAKRPITGASFKTIFDEADALTQDAQALRTMEM 136
QY 150 YAKSISHVILSKTVKGTQKLDSSY-----DALIKEVANGDYIYEANS GAVKRV 203
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 137 FSSNREFILSCYSSKTIETPIQSRCAIFRFRPLDEDIARLR-----YIAENGL 188
QY 204 GRCDSPATEYDLBAEEVYVPIKGEVHKKEIVQDYLHD-----LDAANQPOGGQ 254
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 189 -----LPEEGQALILY--IAEGDMRRAINILQAAALDKKITDENVFMASSRRPDIR 240
QY 255 D--ILSLMGQMKPKRTETETKLRQINKVYNYIDEGIAELVPGVLFIDEVHMLDIECF 312
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 241 EMMLALAKGNFLKAR-----EKRLRELK-----QGLS-----GEVIVGMK---EVF 281
QY 313 SYLNRALSPILSPVILATNRGICNVGRTMTSPHGIVDLRLVITRTYPTQMIO 372
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 282 ---NLPIDEPKK--VLLAKKIGEYNERLVE-----GANETIQ 313
QY 373 ILAIRAQ 379
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 314 LEALLAQ 320

```

```

RESULT 5
US-08-987-123-2
; Sequence 2, Application US/08987123C
; Patent No. 6136557
;
; GENERAL INFORMATION:
; APPLICANT: Eli Lilly and Company
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE GENE SEQUENCE FISH
; FILE REFERENCE: X-11762 Sequence Listing
; PATENT NO. 6136557
; CURRENT APPLICATION NUMBER: US/08/987,123C
; CURRENT FILING DATE: 1997-12-08
; PRIOR APPLICATION NUMBER: 60/036,281
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
;
; US-08-987-123-2

```

```

Query Match          4.9%; Score 111.5; DB 4; Length 652;
Best Local Similarity 23.1%; Pred. No. 0.05;
Matches 112; Conservative 73; Mismatches 162; Indels 137; Gaps 31;

```

```

QY 39 FVGGAAAREAGLAVDMT-----ROKKMAGR---AVLAGPPATGKTALALGIAOELG 88

```

```

? TELEFAX: (617) 876-5851
? INFORMATION FOR SEQ ID NO: 12:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 686 amino acids
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-08-993-228-12

Query Match          4.8%; Score 108.5; DB 2; Length 686;
Best Local Similarity 20.1%; Pred. No. 0.1;
Matches 109; Conservative 74; Mismatches 184; Indels 175; Gaps 25

QY      23 GLGLDANGMATAIAGFGOAA-----REAAGLA-----YDMIRKK-- 60
       ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      185 GIGTGGMGGLS---VGFTTAKVLKDEIDVDFKVAVGCEAKLEIMEVFLLKNPKGY 241
QY      61 -----MAGRAVLLAGPPARGKATALAGIAOELGSKVPFCPMWSEYSSSEVKTEVIMEN 115
       ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      242 QDLGAIIIPKAILTGPGPGKTLLAKATAGE--ANVPFITVSSE-----FILEM 288
QY      116 FRRATGLRIKE-----NKE-----VYGEVETLSPEAEESTGTGYAK----- 152
       | :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      289 FVGVGARVADFLARKNAPCILFTIDTAVGRRKRGNFGGSGEQENTLNLQLVEMDG 348
QY      153 --SISHVIISLKT-----VKG-----TKOLKDSITYDALIKEKVAV-----G 188
       :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      349 FNTTTNVILLAGINRPGRPDIKGRASIFKVVHLRPLKLDSTLEKDKLARKLASLTGPFSGA 408
QY      189 DVTYEANSNGAKVRRCRCSFATEVDLEEEVYVPIPKGVHKKKITVQ----DVTLHDL 243
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      409 DVANNQNEALLIAARHLSDSIHQKHFEQAIERV---IGLKKKQTQLPPEEKTVAYHEA 465
QY      244 DAANO-----POG-----GDQLSLMGOMMPRKTE- 270
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      466 GAHVAGMYLEHADPLLKYSIIPRGKGLGYAOYLPREQYLYTRKEOLLDRMCMTLGGRVSEE 525
QY      271 -----ITEKLROEIKKVVN---RYIDEGIAELYPGVLFIDEVHMLDIETCSYLNRALES 321
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      526 IFFGRITTAGOADDLKKVQSAAVQIVQFMENEKV-GQISFDLPGRGDM-----VLEK 576
QY      322 PLSPVILIATNNGCIGNVAGDTMTSPHGIRPVOLLDR-----LVIRTETPTETMIQ 372
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      577 PYSE-----ATARLIDDEVERRILLINDAYKRTVALITEKKADVERKALLLEKVEVLDKNDIVE 632
QY      373 ILAIARQVEEIDMDESIAYIGEIQOOTSILRHAIOLISPASVSVSKTNGREKICKADEEV 432
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      633 LLGPREFAEK-STEEEFVEGTGSLDEDTSL-----PEBLKOMNNERERK-----EKEEP 679
QY      433 SG 434
       |
Db      680 PG 681

RESULT 7
US-08-928-213B-11
: Sequence 11; Application US/08928213B
: Patent No. 6238905
GENERAL INFORMATION:
APPLICANT: McHenry, Charles S.
           Seville, Mark
           Cull, Millard G.
TITLE OF INVENTION: NOVEL "THERMOPHILIC POLYMERASE III
NUMBER OF SEQUENCES: 195
CORRESPONDENCE ADDRESS:
ADDRESSSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
```


RESULT 9
US-09-388-774-1
; Sequence 1, Application US/09388774
; Patent No. 6228991
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Paterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR

NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Inocyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 OPERATING SYSTEM: DOS
 SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/388,774
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/074,579
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Cerrone, Michael C
 REGISTRATION NUMBER: 39,132
 REFERENCE/DOCKET NUMBER: PF-0505 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-855-0555
 TELEFAX: 650-845-4166
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 942 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: UTRSNOT02
 CLONE: 688183
 US-09-388-774-1

Query Match 4.6%; Score 104.5; DB 4; Length 942;
 Best Local Similarity 18.0%; Pred. No. 0.41;
 Matches 87; Conservative 70; Mismatches 132; Indels 195; Gaps 23;

93 PCPMWSEYSSSEVKTEYLMENFRALGLRIKENK-----EYEGCVTELSP 140
 98 FTMLIGDKVYGEITERE-----KKGDRVKEKRNKTEENGEKTEIFRASAVIPSK 150
 141 EEA-----ESTGGYAKSISHVLIISLKYKGTOKLDDSIYDA----- 179
 151 DKAAFFLSEELORRLKYEHS-----ISVRPOOLSGRLSDVNLIESAGIASLEVLPL 205
 180 -----LIKEVAVGVVY--TEANSNAVRRGRDSDATEYDE 216
 206 HNSRORGSGEDDGPPEPVIINQNETFANIIFFKPTVVOQARIADGILGDFIINYDVA 265
 217 AEEVY-----PIPKG-----EYHKKKEIVODVTLHD 242
 266 REQSIGIOVLNGFYVYFAPKDLPLPKNVVFLDSASWGTAKLRQKDALFTI-LHD 324
 243 LDANAPOGOGDILSLMG-----QMMKPRKTEITEK 274
 325 L-----RP-----QDFSLIGSNRIKWKDHLISVTPDSIRDKVYIHHMSPGTGIDINCA 376
 275 LROINKVNVKRI-DEGIAE-----LVPGVLFIDEVHMLDIECFESYLNRALESPLSP 325
 377 LQRAI-RLNKKVYVHSGIGDRSVSLIVFLDCKPTVGETHTLKI-----LNTREARANGQ 430
 326 IVIATNRGICNVGTDMSPHICIPVDLDRLVYIIRTEYGPTEMIQILAIRAOVEETDM 385
 431 VCFITIGI-----NDVDRLEKIST--ENCGILTRVH-----EEDA 467
 386 DEESLAVLGEIGQGSILRAIOL-ISPASVVSKT-----NGREKICKADLEEVSGLY 436

Db 468 GSOLIGFYDEI--RFLSLDIRIDYPPSSVQATKTLFPNFNSSEIT-----IAGKL 518
 QY 437 LPAK 440
 Db 519 VDRK 522

RESULT 10
 US-08-928-213B-10
 Sequence 10, Application US/08928213B
 Patent No. 6238905
 GENERAL INFORMATION:
 APPLICANT: McHenry, Charles S.
 Seville, Mark
 Cull, Millard G.
 TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III
 HOLOENZYME
 NUMBER OF SEQUENCES: 195
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: MEDLEN & CARROLL, LLP
 STREET: 220 Montgomery Street, Suite 2200
 CITY: San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/928,213B
 FILING DATE: 12-Sep-1997
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: MacKnight, Kamrin T.
 REGISTRATION NUMBER: 38,230
 REFERENCE/DOCKET NUMBER: ENZYCO-02550
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-705-8410
 TELEFAX: 415-397-8338
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 273 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:
 US-08-928-213B-10

Query Match 4.5%; Score 101.5; DB 4; Length 273;
 Best Local Similarity 25.6%; Pred. No. 0.11;
 Matches 55; Conservative 24; Mismatches 73; Indels 63; Gaps 10;

224 PKGEVHKKKEIYVD--VTLHDLDANAPOGOGDILSLMGQMMKPRKTEITEKLRQEI 281
 79 PCGYCDNCRLEIGREYVDLLEIDASR-----TKVEDTRDLIDNV-- 118
 282 VVNRVIDEGIAELVPG-----VLFIDEVHMLDIECFESY-LNRALESPLSP--YIATN 332
 119 -----QVAPARGRFKYVLIDEVHMLSRHSFNKALLKTLDEPXXPEHVFILAT- 166
 333 RGICNVGTDMSPHICIPVDLDRLVYIIRTEYGPTEMIQILAIRAOVEETDMDESLAY 392
 167 -----TDPOKLPVITLIR--CLQFHXKXKALKALDVEQIRHOLEHI-LNEEHIAH 212
 393 LGEIGQ-----QTSILRAIOLISPAVSYSKTNCR 421
 213 EPRAQLIARAAGSLRALSLTDQATASGXDDQ 247

```
RESULT 11
US-09-541-782-2
; Sequence 2, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Berand, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1184
; TYPE: PROT
; ORGANISM: A. nidulans
US-09-541-782-2

Query Match          4.4%; Score 100.5; DB 4; Length 1184;
Best Local Similarity 18.3%; Pred. No. 1.4;
Matches 75; Conservative 81; Mismatches 167; Indels 87; Gaps 17;

OY 86 ELGSKVPPCPVGVSEVSVSEYKTEVLMENFRRAIGRIKENREV-YEGEVTELSPEEAE 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 421 QINSTPMKMTLLREFTAIEKELKALLATRRHNGVYMSVEYEMKRNESRLISEEOR 480
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 145 STTGVAKSISHVILSKTVKGT-KOLKLDSSIVDALIKEKAVAGVI---YIANSKA 199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 481 AIESMESLIRKIVQELLTYSKFNLDKKDN---DPTLAALCSTNDVLOOTDYLQNTRA 537
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 200 ---VRVRCDSFATEYVLEAEYVPIPKGEVHKKEIVQDVT---LHLDLDAANA 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 538 QLEEEEMLCACHEETEHOIQ---DVGKGLITGLGVEDINSLOSKLDKRAELDFTNA 592
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 249 QVGGGDLISLGMKMKPKTEITEKLRQELINKVNRV-----IDEGIAELVPGV 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 593 E-----LMRASSTFVSVDYTRIDQRYEAFQTRAKLLETTSVKNFEIATEISNI 642
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 299 LFIDEVHMLDIECFSYLNPALESPLSPYIILATNNGICNVAGTDMTSPHGIPVLDRLV 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 643 ---ETRSDL---SEYNSRLDA-----ACNNKAETSSAHEDMNVLEIK 682
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 359 IIRTEYVGT--EMIOILIRAOVEFIDMDESLAYLGEIGQOTSLRA-----IQL 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 683 DLKEEVKSKVGEGLNGLSAAARISF-----EVIGFTQLHSOLHTSFNNLGKDLKS 734
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 409 I--SPASVSKTNGREKICKADLEEVSGLYDA--KSSARLLQEOOEVI 454
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 735 IFETMATHLSEOKNEINRLRAELQSSNRONIFETHKASAHLAQALEEHV 784
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
```

```
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTD-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7100
; TELEFAX: (617) 248-7000
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-390-4
```

```
Query Match          4.3%; Score 98; DB 1; Length 2101;
Best Local Similarity 20.7%; Pred. No. 6.1;
Matches 107; Conservative 88; Mismatches 190; Indels 132; Gaps 28;

OY 4 EEVOSTSKORIRATHTHIKIGLDANGMALAAGFVGOAARE-----AAGL 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1340 EQALSTLOLEHTSTQALVSEL-LPAKHLCOOLAQ---EQAAAEKRHRELEQSKQAAGL 1395
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 52 AVMDIRKKMGRAVLALAGPATGKTALALGIAQL-GSKVPPCPVGVSEVSEYKTE 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1396 RAELLRQAQELGELI-----PLRQKVAEQERTQAQLAEKASVAEOL---SMKRAH 1444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 111 VLMEFRRAIGRIKENVEYGE-----VIELS-----PEAEEST--- 146
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1445 GLAEENRGIGERANLQGFLEVLDAQREKYVOELAAVRADAEFLAEVOREAOSTARE 1504
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 147 ---TGVYAKSISHVITISLKTVKGKOLKLDSSIVDALIKEKAVAGVIYIANSKAVKR 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1505 LEVMTAKYEGAAVVKVLEEQRYQREERQ-KITAOVEE--LSKKLSDS---QASKVQOQK 1557
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 203 VGRCSFATEYVLEAEY---VPIPKGEVHKKEIVQDVTILH-----DIDAANAQVGG 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1558 LKAVQAQGESEGOEAFQAOQLNELOAQLSQKEQAHEHYKLQMEKAKTHYDAKQAOQL 1617
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 254 QDILSLGMKMKPKTEI---TEKLRQELINKV-----VNRVIDEGIAELVPGVLET 301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1618 QEOQLRSLEQLQENK-ELNAEAERLGHLEQAGLTKTEAEQTCRLHTAOVRSLEAOVANA 1676
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 302 DEVMHMLDIECFSYLNPALES--PLSPYIILATNNGICNVAGTDMTSPHGIPVLDRLV 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1677 DQ-QLRDLGCFQVATDALASREPOAKPQDLT-----SIDSLDSCBEGPPLSTSKLP 1728
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 359 IIRTEYTG---PEMIOILAIR--AOVEEIDM-----DEBSLAYLGEI--- 396
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1729 --RTQPDGTSVGPSPASPLISQRLPKVSELSIYFTPIPARSOAPLESSLDSLGGVFLDS 1786
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 397 GQGT--SLRHAQILISPAVSKTNGREKICKADLEE 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1787 GRKTRSAARRTQIIN---ITWTK-----KIDVBE 1813
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-08-470-950-4
; Sequence 4, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
```

```

; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,950
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-950-4

```

```

Query Match      4.3%; Score 98; DB 1; Length 2101;
Best Local Similarity 20.7%; Pred. No. 6.1;
Matches 107; Conservative 88; Mismatches 190; Indels 132; Gaps 28;

```

```

QY 4 EVOVSTSKKRIATHTHIGKGLDANGMAIALAGFYGOAARE-----AAGL 51
DB 1340 EOMLSTLQLEHTSTQALVSEL-LPAKHLCQOLA---EQAAEKRRHELEQSKQAAGI 1395
QY 52 AVDMIRKMKAGRAVLLAGPRATGTALALGIAQEL-GSKVPCPMGSEVSEVKKTE 110
DB 1396 RAELRAQRELGLI-----PLRQKVAEQERTAOOLRAEKASYAEOL-----SMKKAH 1444
QY 111 VLMEFRRAIGIRIKENKEVEGE-----VTELS-----PEEAEST--- 146
DB 1445 GLIAEENRGLGERANLGRQFLEVELDQAREKYVDELAANVADAEETRLAEVOREAOSTARE 1504
QY 147 ----TGGAKSISHYIISLTKVGTQKLDSSIYDALIKEKVAAGDVITYIANGSAVKR 202
DB 1505 LEVMTAKTEGAKVAVLEERQFQERQ-KLTAQVEE--LSKKLADSD---QASKVQOQK 1557
QY 203 VGRCDSPATEYDLEAEFY---VPIPKGEVHKKEIYQDVTLLH-----DLDAANQPOGG 253
DB 1558 LKAVOAGGSGEQEQRQAOLNELQALQSOKEQAAEHYKLOMEKAKTHYDAKKQOQNEL 1617
QY 254 ODILSLMGOMKPKRTET---TEKLROEINKV-----VNRVIDEGIAELVPGVLFI 301
DB 1618 QEOURLSEOLQKQENK-ELRAEAERLGHLEQOAGLKTKEAQCIRHLTAQVRSLEAQVANA 1676
QY 302 DEVHMLDIECFSYLNRALS---PLSPIVILATNRGICNVGCDMTSPHGIPVLLDRLV 358
DB 1677 DQ-QLRDLGRQVATDALKSREPOAKPOLDL-----SIDSIDLSCEGEGTPLSTSKLP 1728
QY 359 IIRREYTG---PTEMIOILAIR--AQVEEIDM-----DEESLAVLGEI----- 396
DB 1729 --RTOPDGTSPVGPBPSPISQRLPKVESLESILYTPIPARSCAPLESLSLGDVFLDS 1786
QY 397 GQOFT--SLRAIALQILSPASVSKTNGREKIKADLEE 431
DB 1787 GKRTSRARRRTTOIIN---ITWTK-----KLDAVE 1813

```

```

RESULT 14
US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,781
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-781-4

```

```

Query Match      4.3%; Score 98; DB 1; Length 2101;
Best Local Similarity 20.7%; Pred. No. 6.1;
Matches 107; Conservative 88; Mismatches 190; Indels 132; Gaps 28;

```

```

QY 4 EVOVSTSKKRIATHTHIGKGLDANGMAIALAGFYGOAARE-----AAGL 51
DB 1340 EOMLSTLQLEHTSTQALVSEL-LPAKHLCQOLA---EQAAEKRRHELEQSKQAAGI 1395
QY 52 AVDMIRKMKAGRAVLLAGPRATGTALALGIAQEL-GSKVPCPMGSEVSEVKKTE 110
DB 1396 RAELRAQRELGLI-----PLRQKVAEQERTAOOLRAEKASYAEOL-----SMKKAH 1444
QY 111 VLMEFRRAIGIRIKENKEVEGE-----VTELS-----PEEAEST--- 146
DB 1445 GLIAEENRGLGERANLGRQFLEVELDQAREKYVDELAANVADAEETRLAEVOREAOSTARE 1504
QY 147 ----TGGAKSISHYIISLTKVGTQKLDSSIYDALIKEKVAAGDVITYIANGSAVKR 202
DB 1505 LEVMTAKTEGAKVAVLEERQFQERQ-KLTAQVEE--LSKKLADSD---QASKVQOQK 1557
QY 203 VGRCDSPATEYDLEAEFY---VPIPKGEVHKKEIYQDVTLLH-----DLDAANQPOGG 253
DB 1558 LKAVOAGGSGEQEQRQAOLNELQALQSOKEQAAEHYKLOMEKAKTHYDAKKQOQNEL 1617
QY 254 ODILSLMGOMKPKRTET---TEKLROEINKV-----VNRVIDEGIAELVPGVLFI 301
DB 1618 QEOURLSEOLQKQENK-ELRAEAERLGHLEQOAGLKTKEAQCIRHLTAQVRSLEAQVANA 1676
QY 302 DEVHMLDIECFSYLNRALS---PLSPIVILATNRGICNVGCDMTSPHGIPVLLDRLV 358

```

Db 1677 DQ-QLRDLGKFOVATDALKSREPQAKPOLDL-----SIDSLDSCBEGTPLSTSKLP 1728
 QY 359 IIRFETYG---PTEMIQILAIR--AOVEIDM-----DESLAYIGE1---- 396
 Db 1729 --RTQPDGTSVGPBPASPIQSRLPPKVESLFTPIPARSQAPLBSLSDSLDGVFLDS 1786
 QY 397 GOOT--SLRHA1QILISPAVSCKTNGREKICKADLEE 431
 Db 1787 GRKTRSAARRRTQIIN-----ITMTK-----KLDFEE 1813

RESULT 15
 US-08-483-924-4
 ; Sequence 4, Application US/08483924
 ; Patent No. 5882876
 ; GENERAL INFORMATION:
 ; APPLICANT: TOLKATLY, GARY
 ; APPLICANT: LIDGARD, GRAHAM P
 ; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
 ; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
 ; STREET: 125 HIGH STREET
 ; CITY: BOSTON
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/483,924
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: PITCHER ESQ, EDMUND R
 ; REGISTRATION NUMBER: 27,829
 ; REFERENCE/DOCKET NUMBER: MTP-013
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 248-7000
 ; TELEFAX: (617) 248-7100
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2101 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-483-924-4

Query Match 4.3%; Score 98; DB 2; Length 2101;
 Best Local Similarity 20.7%; Pred. No. 6.1;
 Matches 107; Conservative 88; Mismatches 190; Indels 132; Gaps 28;

QY 4 EEVOSTSKKRIATHTHIKGLIDANGMAIALAAGVGOAARE-----AAGL 51
 Db 1340 EQALSTLQLEHTSTQALVSEL-LPAKHLCQQLQA---EQAAAEKRRRELDOSKQAAGL 1395
 QY 52 AYDMIRQKKMAGRAVLGAPATGKTALALGIAOEL-GSKVPFCPMGSEVYSSSEVKTE 110
 Db 1396 RAELLRAQRELELTI-----PLROKVAQERTAOQLRAEKASYAEOL-----SWLKKAH 1444
 QY 111 VMENFRRAISGLRIKENKEVEGE-----VPELS-----PEEAEST--- 146
 Db 1445 GLLAEENGGLGRANGLQFLEVEDQAREKYVDELAAVRADATETRLAEVQREASTARE 1504
 QY 147 ---TGYAKSISHVITSLKTVKGTQKLDLSDSIYDALIKEKVAAGDVITYEANSQAVKR 202
 Db 1505 LEVMTAKYEGAKVAVLEERQRFQERQ-KLTAQYEE--LSKKIADSD-----QASKVQOQK 1557

QY 203 VGRCDSPATEYVDLEAEY---VPIKGEVHKKEIYQDVTIL-----DLDAANAQPOGG 253
 Db 1558 LKAVQAQGESQOEAORFOAQLNELQAQLSQKEQAAEHYKLMKEKAKTHYDAKKQOQNEL 1617
 QY 254 ODLISLMGOMKPRKTEI---TEKLRQELNKV-----VNRVIDEGLAEVGVFLI 301
 Db 1618 QEOQLRSLEQLOKENEK-ELRAEAERLGHLEQAGLKTKEAEOGTCRHILTAQVRSLEAQVANA 1676
 QY 302 DEVMHMDIECFSYLNRALES---PLSPIVILATNRGICVNRGDMTSPHIGIVDLDRLY 358
 Db 1677 DQ-QLRDLGKFOVATDALKSREPQAKPOLDL-----SIDSLDSCBEGTPLSTSKLP 1728
 QY 359 IIRFETYG---PTEMIQILAIR--AOVEIDM-----DESLAYIGE1---- 396
 Db 1729 --RTQPDGTSVGPBPASPIQSRLPPKVESLFTPIPARSQAPLBSLSDSLDGVFLDS 1786
 QY 397 GOOT--SLRHA1QILISPAVSCKTNGREKICKADLEE 431
 Db 1787 GRKTRSAARRRTQIIN-----ITMTK-----KLDFEE 1813

Search completed: November 13, 2001, 07:01:18
 Job time: 324 sec

OM of: US-09-589-510-4 to: N_Geneseq_0601:* out_format : pfs
Date: Nov 13, 2001 7:57 AM

About: Results were produced by the gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+2pn.model -DEV=xlip
-Q=/cgn2.1/USPTO.spool/US09589510/runat_13112001_065634_13480/app-query.fasta.1.518
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=std.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FAPOP=6.000 -FAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=200000000
-USER=US09589510.@CGN1_1.228 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLIPX
-WAIT -THREADS=1

Search information block:

Query: US-09-589-510-4
Query length: 455
Database: N_Geneseq_0601:*
Database sequences: 730101
Search time (sec): 313950809
Search time (sec): 117.850000

Score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
/SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAAD2567 +			2263.00	3983.19	6.8e-214
/SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAAD2566 +			2253.00	3965.86	6.3e-213
/SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAAD2568 +			2222.00	3910.69	7.4e-210
/SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAAD2569 +			2211.50	3892.02	8.1e-209
/SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAAD2570 +			2157.50	3796.51	1.7e-203
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAC8992 +			1994.00	3508.42	1.9e-187
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAAI8199 +			1723.50	3028.37	1.0e-160
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAAI8198 +			1722.50	3024.74	1.2e-160
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAAX2405 +			1721.00	3027.51	1.7e-160
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAC5922 +			1570.00	2756.31	1.5e-145
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ52915 +			1507.50	2647.14	1.8e-139
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAF3700 +			1064.50	1864.59	6.9e-96
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAF3700 +			973.00	1700.23	9.9e-87
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAF3700 +			932.50	1629.03	9.1e-83
/SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAAF4507 +			676.00	1095.04	5.0e-53
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAAX3016 +			538.50	934.68	4.3e-44
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAF0752 +			525.50	914.80	5.5e-43
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAC0209 +			321.00	558.50	3.9e-23
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAF0804 +			309.00	539.92	4.2e-22
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAC0403 +			308.00	535.37	4.2e-22
/SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT:AAAC0576 +			251.00	437.89	2.0e-16
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAAV2120 +			167.50	196.99	0.0053
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAC4242 +			135.50	213.79	0.0066
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAB1463 +			134.50	171.57	0.1382
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAF2160 +			134.50	155.42	1.10
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAB2660 +			132.50	205.50	0.0018
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAAF4507 +			130.50	210.13	0.0010
/SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT:AAAT2063 +			130.00	129.53	30.36
/SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT:AAAC0559 +			127.00	187.96	0.0169
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAC5155 +			126.00	194.13	0.0077
/SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAAV2287 +			126.00	161.82	0.0362
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAB1501 +			126.00	182.02	0.4822
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAAF2161 +			126.00	140.36	7.56
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAAZ3100 +			125.00	196.32	0.0058
/SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAAF5473 +			125.00	196.32	0.0058
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAAI3165 +			123.50	177.02	0.0687
/SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT:AAAT0996 +			122.00	173.88	0.1027
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ54007 +			121.50	194.10	0.0077

/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ54008 +	121.50	194.10	0.0077	1176
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ56953 +	121.00	167.58	0.0177	1978
/SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ09525 +	121.00	166.32	0.2708	1407
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ54006 +	120.50	192.33	0.0096	1176
/SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ54004 +	120.50	186.95	0.0192	1932
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAAD2567				
seq_documentation_block:				
ID	AAAD2567	standard; cDNA; 1912 BP.		
AC	AAAD2567;			
XX				
DT	02-MAY-2001	(first entry)		
DE	Maize Ruvb orthologue #2 cDNA.			
XX				
KW	Maize; Ruvb orthologue; branch migration; heteroduplex extension; homologous recombination; transformation; transgenic plant; ss.			
XX				
OS	Zea mays.			
XX				
FT	Key	Location/Qualifiers		
FT	CDS	94..1461		
FT	P-PSDB: AAY72563.			
FT	/product= "Maize Ruvb orthologue protein #2"			
PN	W0200105975-A1.			
PD	25-JAN-2001.			
XX				
PR	16-JUL-1999; 99US-0144112.			
XX				
PA	(PION-) PIONEER HI-BRED INT INC.			
XX				
XX	Mahajan PB;			
XX	WPI; 2001-159537/16.			
PT	Ruvb in plants, as probes or amplification primers in the detection, quantitation or isolation of gene transcripts -			
PT	Claim 1; Page 70-72; 87pp; English.			
XX	The present sequence is a Zea mays RuVB orthologue #2 cDNA. RuVB along with RuVA catalyses the branch migration process, also known as heteroduplex extension, in homologous recombination. RuVB is used for the control of homologous recombination or transformation efficiency in transgenic plants. The RuVB nucleotide may be used as probes or amplification primers for detecting, quantifying or isolating gene transcripts, in detecting deficiencies in the mRNA level during screening for desired transgenic plants, for detecting gene mutations, for monitoring upregulation of expression or changes in enzyme activity, for detecting any number of allelic variants, orthologues or paralogues of the gene, or for site directed mutagenesis in eukaryotic cells. It may also be used for recombinant expression of its encoded polypeptide, or for use as immunogen in preparing and/or screening of antibodies, and in sense or antisense suppression of one or more genes in a host cell, tissue or plant. The RuVB proteins may be used in assays to agonise or antagonise the enzyme function, or as immunogens or antigens for screening antibodies.			
XX	Sequence 1912 BP; 547 A; 385 C; 515 G; 462 T; 3 other;			
XX	alignment_scores:			
	Quality: 2263.00	Length: 455		
	Ratio: 4.974	Gaps: 0		
	Percent Similarity: 100.000	Percent Identity: 100.000		

alignment_block:
US-09-589-510-4 x AAD02567

Align seg 1/1 to: AAD02567 from: 1 to: 1912

```

1 MetArg1IleGluValGlnSerThrSerLysLysGlnArg1IleAlaThr 17
  |||
  |||
  |||
94 ATGAGATCGAGGAGGTGCAGTGCAGCTCGAAGACAGCGCATCCGCAC 143
  |||
  |||
  |||
17 RHisThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleA 34
  |||
  |||
  |||
144 CCACACCCACATCAAGGAGCTCGCGCTCGACGCCAATGGATGGCCGATTG 193
  |||
  |||
  |||
34 LalaValAlaIleGlyPheValGlyGlnAlaAlaIleArgGluAlaIleGly 50
  |||
  |||
  |||
194 CGTTGGCGGCGGGGTTCGTGGGCCAGCGCGCGCGCGCGCGCGGG 243
  |||
  |||
  |||
51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaValLe 67
  |||
  |||
  |||
244 CTGGCGGTGCAGCATGTCGCCAGACAGAGATGGCCGCCGCCGGGTGCT 293
  |||
  |||
  |||
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
  |||
  |||
  |||
294 CTTGGCGGTCCGCCCGCCAGCGCGCTAGCGCTCGCATAG 343
  |||
  |||
  |||
84 IArgGlnLeuGlySerLysValProPheCysPrometValGlySerGlu 100
  |||
  |||
  |||
344 CCCAGGAGCTCGCGCAGCAGAGTCCCTTCTGCTATGTAAGTACGAA 393
  |||
  |||
  |||
101 ValTyrSerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
  |||
  |||
  |||
394 GTGTACTCTCGAGGTCAAGAAAACGAGTGTGATGAAATTTCCG 443
  |||
  |||
  |||
117 GArgAlaIleGlyLeuArgGlyLysGluAsnLysGluValTyrGluGly 134
  |||
  |||
  |||
444 TAGAGCTATAGGTTGGCTATAAAGAAAACAAAGAGTTATGAAGAG 493
  |||
  |||
  |||
134 LuValThrGluLeuSerProGluGluValGlnSerThrThrGlyTyr 150
  |||
  |||
  |||
494 AGGTACTGAACTTCCCGAAGAGAGCTGAGACTCACTGTAAAGGAG 543
  |||
  |||
  |||
151 AAlaLysSerIleSerHisValIleIleSerLeuLysThrValLysGly 167
  |||
  |||
  |||
544 GCAAAAAGCATTAACCATTAATCATCAGCTTAAGACTGTAAAGGAG 593
  |||
  |||
  |||
167 rLysGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGlu 184
  |||
  |||
  |||
594 TAGCAACTGAAGTTAGATTCTTCAATTATGATGCTGATCAAGGAAA 643
  |||
  |||
  |||
184 ySValAlaValGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
  |||
  |||
  |||
644 AGGTGGCAGTGGGTGATGTTATATACATGCAACAAATAGTGCAGCAG 693
  |||
  |||
  |||
201 LysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGlu 217
  |||
  |||
  |||
694 AAAAGAGTTGCTAGATGTCATCTTTGCTACAGAAATACATCTTGAAG 743
  |||
  |||
  |||
217 aGluGluTyrValProIleProLysGlyGluValHisLysLysLysGlu 234
  |||
  |||
  |||
744 TGAAGACTAGTTCTTATCCCAAGGTGAAGTCATAGAAAGAAAAA 793
  |||
  |||
  |||
234 lValAlaGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnPro 250
  |||
  |||
  |||
794 TTGTGGCAGGTGTCACACTTCATGACCTTGATGTCACAAATGCTCAGCA 843
  |||
  |||
  |||
251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLeuLysProAr 267
  |||
  |||
  |||
844 CAAGGTGGCCAAGATATTTGTCCCTTATGGGCCAATGATGTAACACAG 893
  |||
  |||
  |||
267 gLysThrGluIleThrGluLysLeuArgGlnGluIleAsnLysValAla 284
  |||
  |||
  |||
894 AAAGACTGAATCACCAGAAAACCTACGCCAAGAAATTAATAGGTGTAA 943
  |||
  |||
  |||

```

```

284 snArgTyrIleAspGluGlyIleAlaGluLeuValProGlyValLeuPhe 300
  |||
  |||
  |||
944 ATGATATATTCGATGAAGCAATTCAGAGCTGTACCTGGTGTTC 993
  |||
  |||
  |||
301 IleAspGluValHisMetLeuAspIleGlyCysPheSerTyrLeuAsnAr 317
  |||
  |||
  |||
994 ATTGATGAGTCCACATGTTGGATATCGAATGTTTCTTATCTTAACCG 1043
  |||
  |||
  |||
317 gAlaLeuGluSerProLeuSerProIleValIleLeuAlaThrAsnArg 334
  |||
  |||
  |||
1044 TGCATTGGAGAGCCCAATTATCCAAATCGTGAATCTGCTACAAATAGCG 1093
  |||
  |||
  |||
334 lYleCysAsnValArgGlyThrAspMetThrSerProHisGlyIlePro 350
  |||
  |||
  |||
1094 GAAATATGTAATGTAAGAGCACTGATGACAAATGCCATGGTATACCG 1143
  |||
  |||
  |||
351 ValAspLeuLeuAspArgLeuValIleIleArgThrGluThrTyrGlyPr 367
  |||
  |||
  |||
1144 GTGGATCTTCTAGATAGGCTGGTGAATTTATTCGACAGAGACATATGGCC 1193
  |||
  |||
  |||
367 oThrGluMetIleGlnIleLeuAlaIleArgAlaGlnValGluGluIleA 384
  |||
  |||
  |||
1194 TACTGAGATGATACAGATATTGCTATCCGACACAAAGTGGAGGAGATTG 1243
  |||
  |||
  |||
384 sPMeTAspGluGluSerLeuAlaTyrLeuGlyGluIleGlyGlnGlnThr 400
  |||
  |||
  |||
1244 ATATGATGAAGAAAGCTCTTGCTTATTTAGCGGAGATGGAGCAGACACA 1293
  |||
  |||
  |||
401 SerLeuArgHisAlaIleGlnLeuIleSerProAlaSerValSerLys 417
  |||
  |||
  |||
1294 TCTTTGAGACATGCTATTCATTTGATATCACCTCGCACCGGTCTCAA 1343
  |||
  |||
  |||
417 sThrAsnGlyArgGlyLysIleCysLysAlaAspLeuLysGluValSerG 434
  |||
  |||
  |||
1344 GACTATGGAGAGAGAAATCTGCAGAGCTGATCTCAGGAGACACATG 1393
  |||
  |||
  |||
434 lYleuTyrLeuAspAlaLysSerSerAlaArgLeuLeuGlnGlnGln 450
  |||
  |||
  |||
1394 GGCCTATTGATGATGCCAAATCTCGGCTCGGCTGCTCCAGAGACACAA 1443
  |||
  |||
  |||
451 GluArgTyrIleThr 455
  |||
  |||
  |||
1444 GAAAGATACATCACC 1458
  |||
  |||
  |||
seq_name: /SISL/gcdata/geneseq/geneseqn/NA2001.DAT.AAD02566
seq_documentation_block:
ID AAD02566 standard; cDNA; 1845 BP.
XX
AC AAD02566;
XX
DT 02-MAY-2001 (first entry)
XX
DE Maize RuVB orthologue #1 cDNA.
XX
KW Maize; RuVB orthologue; branch migration; heteroduplex extension;
  homologous recombination; transformation; transgenic plant; ss.
XX
OS Zea mays.
XX
FT Key Location/Qualifiers
FT CDS 85..1452 /*tag= a
FT /product= "Maize RuVB orthologue protein #1"
XX
PD 25-JAN-2001.
XX
PF 13-JUN-2000; 2000MO-US16271.
XX
PR 16-JUL-1999; 99US-0144112.

```


XX
PA (PION-) PIONEER HI-BRED INT. INC.
XX
PI Mahajan PB:
XX WPI; 2001-159537/16.
DR P-PSDB; AAY72562.
XX
PT Novel maize Ruvb nucleic acid useful for modulating levels of maize
PT Ruvb in plants, as probes or amplification primers in the detection,
PT quantitation or isolation of gene transcripts
PS
PS Claim 1; Page 67-69; 87pp; English.
CC The present sequence is a Zea mays Ruvb orthologue #1 cDNA. Ruvb along
CC with Ruvb catalyzes the branch migration process, also known as
CC heteroduplex extension, in homologous recombination. Ruvb is used for
CC the control of homologous recombination or transformation efficiency in
CC transgenic plants. The Ruvb nucleotide may be used as probes or
CC amplification primers for detecting, quantifying or isolating gene
CC transcripts, in detecting deficiencies in the mRNA level during screening
CC for desired transgenic plants, for detecting gene mutations, for
CC monitoring upregulation of expression or changes in enzyme activity, for
CC detecting any number of allelic variants, orthologues or paralogues of
CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
CC also be used for recombinant expression of its encoded polypeptide, or
CC for use as immunogen in preparing and/or screening of antibodies, and in
CC sense or antisense suppression of one or more genes in a host cell,
CC tissue or plant. The Ruvb proteins may be used in assays to agonise or
CC antagonise the enzyme function, or as immunogens or antigens for
CC screening antibodies.
XX
SQ Sequence 1845 BP; 530 A; 378 C; 490 G; 445 T; 2 other:

alignment_scores:
Quality: 2253.00 Length: 455
Ratio: 4.952 Gaps: 0
Percent Simlarity: 100.000 Percent Identity: 99.341

alignment_block:
US-09-589-510-4 x AAD02566 ..

Align seg 1/1 to: AAD02566 from: 1 to: 1845

1 MetArlleGlulgluValGlnSerThrSerLysLysGlnArlleAlaTh 17
|||||
85 ATGAGATCGAGGAGGTGCGACCTCGACCTCAAGAACAGCGATCGCCAC 134
17 RHSTHHisIleLysGlyLeuLysLysMetAlaHisngLysMetAlaIleA 34
|||||
135 CCACACCCACATCAAGGACTCGGCTCGACGCCCAATGGATGGCGATTG 184
34 LaleuAlaIaIaGlyPheValGlyGlnAlaAlaIaAargLjuaIaIaGly 50
|||||
185 CGTTGGGGGGGGGTTGTTGGCCAGTGGCGGGCGGCGAGGGCGGGG 234
51 LeuAlaValAspMetIleAargLjuaLysLysMetAlaGlyAArgAlaValle 67
|||||
235 CTGGCGGTGCGACATGATTCGCCAGAAAAMAGATGGCGGGCGCGGTGCT 284
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
|||||
285 CCTTGGGTGCGCCCGCCAGGGCAGAGACGGCGCTRAGCGCTGGCATG 334
84 lAglngLleuGlySerLysValAProPheCysProMetValGlySerGlu 100
|||||
335 CCCAGGACTCGGCGAGGATGCTTCTGCTTATGATGATGATGATGAA 384
101 ValLysSerSerGluValLysLysThrGluValLeuMetGluAsnPhar 117
|||||
385 GTTACTCTCTCGAGGTCAGAAACAGAGAGTGTGATGGAAATTTCCG 434

117 gArGAlaIleGlyLeuAargLjuaLysLysGlnLysGluValAryrGluGlyG 134
|||||
435 TAGAGCTATAGTTGGCTATAAAGAAAAACAAAGAGCTTATGAAGAGG 484
134 luvAlThrGluLeuSerProGluLjuaIaGluSerThrThrGlyGlyTyr 150
|||||
485 AGGTTACTGAACCTTCCCGACAGAGAGCTGAGATACACCTGGGTGATAT 534
151 AlalysSerIleSerHisValIleIleSerLeuLysThrValLysGlyTh 167
|||||
535 GCAAAAGCATTAGCCATGATATCATCGCTTAAGACCTGTTAAAGGAGC 584
167 rLysGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGlu 184
|||||
585 TAGAGCACTGAAGTTAGATCTTCATATTTATGATGCTGTGATCAAGGAA 634
184 ysaValaIaValGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
|||||
635 AGGTGGCAGTGGGTGATCTTATATACATTGAAGCAAAATAGTGGAGCG 684
201 LysAryValGlyAryCysAspSerPheAlaThrGluTyrAspLeuGluAl 217
|||||
685 AAAAGAGTTGGTAGATGTGATCTTTCTTACAGAAATACGATCTTGAAAG 734
735 TGAGAGATGTGTTCCATCCCAAGGTGAAGTCCATGAAGAAAAAGGAA 784
217 aGluGluTyrValProIleProLysGlyGluValHisLysLysGluTyr 234
|||||
234 lAvalGlnAspValThrLeuHisAspLeuAspAlaIaAsnAlaGluPro 250
|||||
785 TAGTGAGAGATGTCAACACTTCACTGATGACGAGCAAAAGCTCAGCCA 834
251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLysProAr 267
|||||
835 CAAGGTGGCCAGATATTTTGTCCCTTANGGCCAGATGATGAAGAACCG 884
267 gLysThrGluIleThrGluLysLeuAargGlnIleLysLysValAla 284
|||||
885 AAAGACTGAATATCCGAAACAACTACGCCCAAGAAATTAATAAGGTGGTA 934
284 snAryrThrIleAspGluGlyIleAlaGluLeuValProGlyValLeuPhe 300
|||||
935 ATGATATATCGATGAAGAAATTCAGAGCTTGACCTGCTGCTTTTATTC 984
301 lIeAspGluValHisMetLeuAspIleGluCysPheSerTyrLeuAsnAr 317
|||||
985 ATTGACAGAGTCCACATGTTGATTCGAATGTGTTTCTTATCTTAACCG 1034
317 gAlaLeuGluSerProLeuSerProIleValIleLeuAlaThrAsnAryG 334
|||||
1035 TGCATTGGAGAGCCCATTTATCACCATTGTGATCTGTGACGAATAGGG 1084
334 lYlIecysAsnValAargGlyThrAspMetThrSerProHisGlyIlePro 350
|||||
1085 GAATATATTAATTAAGAGAACCTGATATGACAAGTCCACATGATGTAACA 1134
351 ValAspLeuLeuAspAargLeuValIleIleAryThrGluThrTyrGlyPyr 367
|||||
1135 GTGATCTTCTAGATAGGTGTGGATATATTCGAGACAGACATATGGCCC 1184
367 cThGluMetIleGlnIleLeuAlaIleAargAlaGlnValAlaGluIleA 384
|||||
1185 TACTGAGATGATACAGATATTTGGCTATCCAGACACAAGTGGAGACATGG 1234
384 spMeLAspGluGluSerLeuAlaTyrLeuGlyIleGlyGlnGlnThr 400
|||||
1235 ATATGATGAAGAAAGCTTGTATATTAGCGAGATCGGACAGCAGACA 1284
401 SerLeuAryHisAlaIleGlnLeuIleSerProAlaSerValAserLys 417
|||||
1285 TCTTTAAGACATCTTATTCATTCATATCATCTGACGCGCTGCTCAAA 1334
417 sThrAsnGlyAryGluLysIleCysLysAlaAspLeuGluValAserG 434

|||||
 1335 GACCTAATGAGAGAGAAATATGCAAGGCTGATCTCGAGGAGTCAGTG 1384
 434 TLeuTYrLeuAspAlaLysSerSerAlaArgLeuGlnGlnGlnGln 450
 1385 GGCCTATTGATGGATGCCAATTCCTGGCTGCTGCCAGGACAAACA 1434
 451 GlnArgTYrIleThr 455
 |||||
 1435 GAAAGATACATCACCC 1449

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:AAD02568

seq_documentation_block:

ID AAD02568 standard; CDNA; 1886 BP.

AC AAD02568;

DT 02-MAY-2001 (first entry)

DE Maize RuvB orthologue #3 CDNA.

KW Maize; RuvB orthologue; branch migration; heteroduplex extension;
 homologous recombination; transformation; transgenic plant; ss.

OS Zea mays.

FT Key Location/Qualifiers

FT CDS 82..1449

FT /tag= a /product= "Maize RuvB orthologue protein #3"

PN W020105975-A1.

PD 25-JAN-2001.

PE 13-JUN-2000; 2000MO-US16271.

PR 16-JUL-1999; 99US-0144112.

XX (PION-) PIONEER HI-BRED INT INC.

XX Mahajan PB;

DR WPI: 2001-159537/16.

DR P-PDB: AAY72564.

PT Novel maize RuvB nucleic acid useful for modulating levels of maize
 RuvB in plants, as probes or amplification primers in the detection,
 quantitation or isolation of gene transcripts -

PS Disclosure: Page 74-76; 87pp: English.

XX The present sequence is a Zea mays RuvB orthologue #3 CDNA. RuvB along
 CC with RuvA catalyses the branch migration process, also known as
 CC heteroduplex extension, in homologous recombination. RuvB is used for
 CC the control of homologous recombination or transformation efficiency in
 CC transgenic plants. The RuvB nucleotide may be used as probes or
 CC amplification primers for detecting, quantifying or isolating gene
 CC transcripts, in detecting deficiencies in the mRNA level during gene
 CC monitoring upregulation of expression or changes in enzyme activity, for
 CC detecting any number of allelic variants, orthologues or paralogues of
 CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
 CC also be used for recombinant expression of its encoded polypeptide, or
 CC for use as immunogen in preparing and/or screening of antibodies, and in
 CC sense or antisense suppression of one or more genes in a host cell,
 CC tissue or plant. The RuvB proteins may be used in assays to agonise or
 CC antagonise the enzyme function, or as immunogens or antigens for
 CC screening antibodies.

SO Sequence 1886 BP; 524 A; 408 C; 518 G; 436 T; 0 other;

alignment_scores:

Quality: 2222.00 Length: 455
 Ratio: 4.916 Gaps: 0
 Percent Similarity: 99.341 Percent Identity: 97.582

alignment_block:

US-09-589-510-4 x AAD02568 ..

Align seg 1/1 to: AAD02568 from: 1 to: 1886

1 MetArgIleGlnGluValGlnSerThrSerLysLysGlnArgIleAlaThr 17
 82 ATGAGGATAGAGAGAGGTGCATGACCTGCAAGAGACGACGATGCCAC 131
 17 rHsThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleA 34
 133 CCACACTCACATCAAGGCCCTCGGCTCGAGCCCAATGGAATGTGATGC 181
 34 LaleuAlaIaGlyPheValGlyGlnAlaAlaAlaArgGlnAlaIaGly 50
 182 CGTTGGCGGGGGGTTGTGGGCGAGGCGCGCGCGAGCGCGCGG 231
 51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaValle 67
 232 CTGGCGGTGCACATGATGATCCGCCAGAGAGATGGCGGTCCGCGGTGCT 281
 67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
 282 CTTGGCGGGCGCGCGCGCGCGCAAAACGGCGCTACCGCTCGGCATAG 331
 84 lAGlnGluLeuGlySerLysValProPheGlyPrometValGlySerGlu 100
 332 CGCAGAGAGCTCGGCGAGCAAGGTCCATTCTGTCTATGAGATCGAGAA 381
 101 ValTYrSerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
 382 GTGTACTCCTCAGAGTCAGAGCAAGAACTGAGGTGCTGATGCAAAATTTCCG 431
 117 gArgAlaIleGlyLeuArgIleLysGluAsnLysGluValTYrGluGlyG 134
 432 TAGAGCTATAGGTTTGGCTATAAAGGAAACAAAGAGTTTATGAAGAGAG 481
 134 lUValThrGluLeuSerProGluGluAlaGlnSerThrThrGlyTYr 150
 482 AGGTTATTGAACCTTCCCGAGAGAGGCTGAGAGACACAACTGGTGATAT 531
 151 AlalysSerIleSerHisValIleIleSerLeuLysThrValLysGlyIle 167
 532 GCGAAAGCATTTAGCCACAGTAATCATTTGGCTTAAAGACTGTCAAGAGGAC 581
 167 lFysGlnLeuLysLeuAspSerSerIleTYrAspAlaLeuIleLysGluL 184
 582 TAAACAATTGAATTAAGACCTTCAATTTATGATGCTTGTGATCAAGGAA 631
 184 ySValAlaValAlaGlyAspValIleTYrIleGluAlaAsnSerGlyAlaVal 200
 632 AGGTGGCAGTGGGTGATGATATATACATTGACCAATAGTGGAGAGAGG 681
 201 LysArgValAlaArgCysAspSerPheAlaThrGluTYrAspLeuGluAl 217
 682 AAAAGAGTTGGTAGATGTATCTTTTCTACAGAAATATGATCTGAAGC 731
 217 aglGluTYrValProIleProLysGlyLysValHisLysLysGluI 234
 732 TGAAGAGTATGCTTATATCCCAAGAGTGAAGTCAATAGAAAGGAA 781
 234 lValGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnPro 250
 782 TAGTCAGAGATGTCACATCTCATGACCTTGATGACGCAAAATGCCAGGCA 831
 251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetMetLysProAr 267

```

832 CAAGGTGGCCAAAGATATTTGTCCCTTATGCGCCAGATGATGATGACCCAG 881
267 gLysThrGluLeuThrGluLysLeuArgGlnGluIleAsnLysValValA 284
882 GAAGACTGAATCACCGAAAGCTACGCCAGAAATCAATAGGTGGTAA 931
284 snArgTyrIleAspGluGlyIleAlaGluLeuValProGlyValLeuPhe 300
932 ACAGTATATGACGAGAGAGATCCAGACGCTTGACCGGTGGTTGGTTC 981
301 IleAspGluValHisMetLeuAspIleGluCysPheSerTyrLeuAsnAr 317
982 ATGTGATGAGCTCCACATGTGTGATATGAAAGCTTTCTTATCTTAACCG 1031
317 gAlaLeuGluSerProLeuSerProIleValIleLeuAlaThrAsnArg 334
1032 TGCATTGGAGAGCCCATTTATCCAAATTTGTATCTGCTACGAAATGAG 1081
334 IlyIleCysAsnValAlaArgIlyThrAspMetThrSerProHisGlyIlePro 350
1082 GAATATGTAATGTGAGAGAGACGATATGACGAGTCCACATGATATACCA 1131
351 ValAspLeuLeuAspArgLeuValIleIleArgThrGluThrTyrGlyPr 367
1132 GTGACCTTCTAGATAGCTGTGATATTTCGACAGAAACATATGAGCCC 1181
367 cThrGluMetIleGlnIleLeuAlaIleArgAlaGlnValGluGluLea 384
1182 TACTGACATGATACGATACTGCTATCCGACGCAAGTGGAAAGAGATTG 1231
384 sPmetAspGluGluSerLeuAlaTyrLeuGlyGluIleGlyGlnGlnThr 400
1232 ATATGATGAGAAAGATCTCTTATTATTAGCGAGATCGGACACACACA 1281
401 SerLeuArgHisAlaIleGlnLeuIleSerProAlaSerValValSerly 417
1282 TCTTTGAACATGCTATTCAGTCTCTATCACCTCCACGCGGTGCGCAA 1331
417 sThrAsnGlyArgGluLysIleCysLysAlaAspLeuGluGluValSerG 434
1332 GACCAACGCGGAGAAAGATGTCAAGGCTGACCTCGAGGAACTCAGCG 1381
434 IlyLeuTyrLeuAspAlaLysSerSerAlaArgLeuGlnGluGln 450
1382 GGCTCTATTGATGCCAAATCCGCGCTCGCTCCAGACACACAA 1431
451 GluArgTyrIleThr 455
1432 GAAAGATACATCACCC 1446
seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.AAD02569
seq_documentation_block:
ID AAD02569 standard; cDNA: 1898 BP.
XX
AC AAD02569;
XX
XX 02-MAY-2001 (first entry)
XX
XX Maize RuVb orthologue #4 cDNA.
XX
XX Maize; RuVb orthologue; branch migration; heteroduplex extension;
XX homologous recombination; transformation; transgenic plant; ss.
XX
XX Zea mays.
XX
XX Key Location/Qualifiers
XX FH 166..1536
XX CDS /*tag= a
XX FT /product= "Maize RuVb orthologue protein #4"
XX
XX WO200105975-A1.

```

```

PD 25-JAN-2001.
XX
XX PF 13-JUN-2000; 2000WO-US16271.
XX
XX PR 16-JUL-1999; 99US-0144112.
XX
XX PA (PION-) PIONEER HI-BRED INT INC.
XX
XX PI Mahajan PB;
XX
XX DR WPT: 2001-159537/16.
XX
XX P-PSDB: AAY72565.
XX
XX PT Novel maize RuVb nucleic acid useful for modulating levels of maize
XX RuVb in plants, as probes or amplification primers in the detection,
XX quantitation or isolation of gene transcripts -
XX
XX PS Disclosure: Page 77-79; 87pp; English.
XX
XX CC The present sequence is a Zea mays RuVb orthologue #4 cDNA. RuVb along
XX with RuVb catalyses the branch migration process, also known as
XX heteroduplex extension, in homologous recombination. RuVb is used for
XX the control of homologous recombination or transformation efficiency in
XX transgenic plants. The RuVb nucleotide may be used as probes or
XX amplification primers for detecting, quantifying or isolating gene
XX transcripts, in detecting deficiencies in the mRNA level during screening
XX for desired transgenic plants, for detecting gene mutations, for
XX monitoring upregulation of expression or changes in enzyme activity, for
XX detecting any number of allelic variants, orthologues or paralogues of
XX the gene, or for site directed mutagenesis in eukaryotic cells. It may
XX also be used for recombinant expression of its encoded polypeptide, or
XX for use as immunogen in preparing and/or screening of antibodies, and in
XX sense or antisense suppression of one or more genes in a host cell,
XX tissue or plant. The RuVb proteins may be used in assays to agonise or
XX antagonise the enzyme function, or as immunogens or antigens for
XX screening antibodies.
XX
XX SQ Sequence 1898 BP; 532 A; 413 C; 521 G; 432 T; 0 other;
XX
XX alignment_scores:
XX quality: 2211.50 length: 456
XX ratio: 4.893 gaps: 1
XX Percent Similarity: 99.123 Percent Identity: 97.368
XX
XX alignment_block:
XX US-09-589-510-4 x AAD02569 ..
XX
XX Align seq 1/1 to: AAD02569 from: 1 to: 1898
XX
XX 1 MetArgIleGluGluValAlaGlnSerThrSerLysLysGlnArgIleAlaThr 17
XX |||||||||||||||||||||||||||||||||||||||||||||||||||
XX 166 ATGAGATAGAGAGAGGTGCATTCGACCTCCAAAGAACACACGATCCAC 215
XX |||||||||||||||||||||||||||||||||||||||||||||||||||
XX 17 rHisThrHisIleLysGlyLeuGlyLeuAsp...AlaAsnGlyMetAlaI 33
XX |||||||||||||||||||||||||||||||||||||||||||||||||||
XX 216 CCACACTCACATCAAGGGCTCGGCTCGACGACGCGCATGGAATTCGA 265
XX |||||||||||||||||||||||||||||||||||||||||||||||||||
XX 33 leAlaLeuAlaAlaGlyPheValGlyGlnAlaAlaAlaArgGluAlaAla 49
XX :: |||||||||||||||||||||||||||||||||||||||||||||||
XX 266 TGCCTGTGGCGGCGGCTTCGTGGCCAGGCGGCGCGCGCGAGCGGCC 315
XX |||||||||||||||||||||||||||||||||||||||||||||||||||
XX 50 GlyLeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaVa 66
XX |||||||||||||||||||||||||||||||||||||||||||||||||||
XX 316 GGGCTGGCGGTGCATCATCCGCCAGAAAGATGCGCGTGGCGCGCT 365
XX |||||||||||||||||||||||||||||||||||||||||||||||||||
XX 66 lleuLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyI 83
XX : |||||||||||||||||||||||||||||||||||||||||||||||
XX 366 GCTCCTTGGCGGCGCGCCGCCAGGCAAAAGCGGTAGCGCTGGCA 415
XX |||||||||||||||||||||||||||||||||||||||||||||||||||
XX 83 leAlaGlnGluLeuGlySerLysValProPheCysProMetValGlySer 99
XX |||||||||||||||||||||||||||||||||||||||||||||||||||
XX 416 TAGCGCAGAGAGCTCGGCGACGAGGTCCCATCTGTCTATGTGTAGATCA 465

```

100	GLVTVTSerSergIuVallyslsYrHgIuValleuMeGluAspH	116
116		
466	GAAGTGTACTCCACAGAGGTCAAGAAACTGAGGTGATGAGAAATTT	515
116	eArGArGAlaIleGlyLeuArGtlleYsgLlAsnYsgIuValYrGluG	133
516		
133	CGGTAGAGCTATGAGTTCGCTGTAAAGGAAAAACAAGAGGTTATGAG	565
133	LyGIuValYrGluLeuSerProGluGluValleuSerThrThrGlyGly	149
566		
150	GAGAGGTTATGAACTTCCCAAGAGAGGTGAAGCAACTGTGTGA	615
150	TYrAllySerIleSerHisValIleIleSerLeuYsThrVallysgI	166
616		
616	TATGCGAAAAGCAATTGACCCAGTAATCAATTGGCTTAAAGACTGTCAAG	665
166	YrHrYsGlnLeuLysLeuAspSerSerIleYrAspAlaLeuIleYsg	183
666		
183	GACTAAGCAATTGCAAAATTGAGCCCTTCAATTTATGATGTTGATCAAG	715
183	LuYsValAlaValAlaYsAspValIleYrIleGluAlaAsSerGlyLa	199
716		
716	AAAAGGTGCAGGTGGGTGATGTTATrACATTGAAGCAATrGTGAGACA	765
200	VallysaYrValGlyYrGysAspSerPheAlaYrHrGluYrAspLeuG	216
766		
216	GTGAAAAGAGTGGTGTGATGTATCTTTTGGCTACAGAAATrAGTCTTGA	815
216	uAlaGluGluYrValProIleProLysGlyLuValHisYsYsYsg	233
816		
816	AGCTGAGAGTATrGTTCCTrATCCCAAGAGTGAAGTCACTTAAGAAAAAG	865
233	LuIleValGlnAspValYrHrLeuHisAspLeuAspAlaAlaAsnAlaGln	249
866		
866	AAATrAGTCAGAGATGTGCACATrCCATrGACTTGTGATGCAAGCAATGCCAG	915
250	ProGlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLysPr	266
916		
916	CCACAAGGTGGCCAAATrTTTTGTCCCTrTAGGGCCAGATrGTAAAGCC	965
266	oArGlySTrHrGluIleThrGluLysLeuArGlnGluIleAsnYsValY	283
966		
966	ACGGAAGACTGAATrACCAGAAAGACTrACGCAAAATrCAATrAAGTGG	1015
283	AlaAsnArGTrYrIleAspGluGlyIleAlaGluLeuValProGlyValLeu	299
1016		
1016	TAAACAGATrATrTCAGCGAAGAAATrCGAGACTTGTACCTGTGGTGTTCG	1065
300	PheIleAspGluValHisMetLeuAspIleGlyCysPheSerYrIleAs	316
1066		
1066	TTCATTGATrGAGGTCCACATrGTGGATrTTGATrTGATrCTTTCTrATTCTAA	1115
316	nArGAlaLeuGluSerProLeuSerProIleValIleLeuAlaThrAsnA	333
1116		
1116	CCGTGATTGAGAGCCCATrTATrCACAATrGTACTGTCTACGTACGATA	1165
333	rGgIlyIleYsAsnValArGlyYrAspMetThrSerProHisGlyLe	349
1166		
1166	GAGGAATrATrGTATrGAGAGGAAACCGATrATrCACAGrTCACATrGTATA	1215
350	ProValAspLeuLeuAspYrGluValIleIleArGTrHrGluYrYrG	366
1216		
1216	CCAGTGTGACTTCTrGATrATrAGGTGGTGTATrTTCGGACAGAAATrATG	1265
366	YrProHrGluMetIleGlnIleLeuAlaIleArGAlaGlnValGluGlu	383
1266		
1266	CCCTACTGTAGATrACAAATrACTGTGTATrCCAGACACAAATrGTAGAAAGA	1315
383	IleAspMetSPrGluGluSerLeuAlaYrIleuGlyGlnIleGlyGlnGln	399
1316		
1316	TTGATrATrGCATrAGAAATrCTTGTGTTrATrTAGGGAATrTGAGACACAG	1365

CC	400	TheRtLeuAArgHisAlaIleuLeuLeuSerProAlaSerValIvalS	416
XX			
XX	1366	ACATCTTGTAGACATGCTATTTCAGTTGGTAACCTGCCACCGTGCGC	1415
XX	416	rLrThraSnGLyArGluLySLecYsLYsALAAsplenuLuInValS	433
XX	:		
XX	1416	AAGAACAACGGGAGAGAAAGATGTGCAGAGCTGCACCTCGAGGAATCA	1465
XX	433	eRGlyLeuTYrLeuASpalalysSerSerAlaArGLEuInGLuIn	449
XX			
XX	1466	GCGGGCTCATTTTGGATGCCAAATCCTGCCCTGCCTCACAGAACAA	1515
XX	450	GIInGUArGYrILeThr	455
XX			
XX	1516	CAGAAAGATACATCACC	1533
XX	seq_name: /SIDSI/gcndata/geneseq/geneseqn/NA2001.DAT:AMD02570		
XX	seq_documentation_block:		
XX	ID	AMD02570 standard; cDNA; 1869 bp.	
XX	AD02570;		
XX	02-MAY-2001	(first entry)	
XX	DE	Maize Ruvb orthologue #5 cDNA.	
XX	KW	Maize; Ruvb orthologue; branch migration; heteroduplex extension; homologous recombination; transformation; transgenic plant; ss.	
XX	OS	Zea mays.	
XX	Key	Location/Qualifiers	
XX	FT	64..1380	
XX	CDS	/*tag= a /product= "Maize Ruvb orthologue protein #5"	
XX	WN020105975-Al.		
XX	25-JAN-2001.		
XX	PD		
XX	PF	13-JUN-2000; 2000WC-US16Z71.	
XX	PR	16-JUL-1999; 99US-0144112.	
XX	PA	(PION-) PIONEER HI-BRED INT INC.	
XX	PI	MahaJan PB;	
XX	DR	WPJ: 2001-159537/16.	
XX	DR	P-PSDB; AAV72566.	
XX	PT	Novel maize Ruvb nucleic acid useful for modulating levels of maize RuvB in plants, as probes or amplification primers in the detection, quantitation or isolation of gene transcripts -	
XX	Claim 1;	Page 80-82; 87pp; English.	
XX	PS		
XX	CC	The present sequence is a Zea mays RuVB orthologue #5 cDNA. RuVB along with Ruva catalyses the branch migration process, also known as heteroduplex extension, in homologous recombination. RuVB is used for the control of homologous recombination or transformation efficiency in transgenic plants. The RuVB nucleotide may be used as probes or amplification primers for detecting, quantifying or isolating gene transcripts, in detecting deficiencies in the mRNA level during screening for desired transgenic plants, for detecting gene mutations, for monitoring upregulation of expression or changes in enzyme activity, for detecting any number of allelic variants, orthologues or paralogues of the gene, or for site directed mutagenesis in eukaryotic cells. It may also be used for recombinant expression of its encoded polypeptide, or for use as immunogen in preparing and/or screening of antibodies, and in sense or antisense suppression of one or more genes in a host cell, tissue or plant. The RuVB proteins may be used in assays to agonise or	

CC antagonise the enzyme function, or as immunogens or antigens for
 screening antibodies.

XX
 SQ Sequence 1869 BP: 539 A; 373 C; 486 G; 471 T; 0 other;

alignment_scores:

Quality: 2157.50 Length: 455
 Ratio: 4.926 Gaps: 2
 Percent Similarity: 96.264 Percent Identity: 96.264

alignment_block:

US-09-589-510-4 x AAD02570 ..

Align seg 1/1 to: AAD02570 from: 1 to: 1869

```

1 MetArgIleGluGluValGlnSerThrSerLysGlnArgIleAlaThr 17
|||||
64 ATGAGGATGAGGAGGTGCGACCTCGAACGACCGCATCGCCAC 113
17 rHsThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleA 34
|||||
114 CCACACCCACATCAAGGACGCTGCGCTCGACGCCAATGGGATGGCATGG 163
34 LaleuAlaIaIaGlyPheValGlyGlnAlaAlaIaArgGluAlaIaGly 50
|||||
164 CGTGGCGGGCGGCTTCGTGGGC..... 186
51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaValle 67
|||||
187 .....CAGAAGAAGATGGCCGCCGCCGCGTGT 215
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
|||||
216 CCTGGCGGGTCCGCCGCCACGGGACAGCGCGCTAGCG...GGCATGG 262
84 lAglnGluLeuGlySerLysValProPheCysProMetValGlySerGlu 100
|||||
263 CCAGAGAGCTCGGACAGCAAGTCCCTTCTGTCGTATGATGATCAGAA 312
101 ValTyrSerSerGluValLysLysThrGluValIleuMetGluAspPheAr 117
|||||
313 GTGTACTCTCGAGGTCAAGAAACTGAGTGTGATGGAATTTCCG 362
117 gArGAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGluGlyG 134
|||||
363 TAGAGCTATAGGTTTGGCTATTAAGGAACAACAAGGTTTATGAAAGAG 412
134 luValThrGluLeuSerProGluGluAlaGluSerThrThrGlyGlyTyr 150
|||||
413 AGGTACTGAACCTTCCCGAAGAGGCTGAGATACAACTGCTGATAT 462
151 AlAlaYsSerIleSerHisValIleIleSerLeuLysThrValLysGlyTh 167
|||||
463 GCAGAAAAGCATTAAGCATATATCAATCAAGCTTAAGACTGTTAAAGGAC 512
167 rLysGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGluL 184
|||||
513 TAGCACTAGCAAGTATGATCTTCAATTATATGATCTCTGATCAAGGAAA 562
184 ysValAlaIaIaGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
|||||
563 AGGTGCAAGTGGTATGTTATATATATCAATCAAGCAATATATGACAGTG 612
201 LysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGluAl 217
|||||
613 AAAAGAGTTGGTAGATGATCTTTTCTCTACAGATATAGATCTTGAAGC 662
217 agLugluTyrValProIleProLysGlyGluValHisLysLysGluI 234
|||||
663 TGAAGGATATGTTCTATCCCAAGAGTGAAGTGCATTAAGAAAAAGAAA 712
234 lValAlaGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnPro 250

```

```

|||||
713 TTGTCAGAGATGTCACACTTCATGACCTTGATGACGCAAAATGCTCAGCA 762
251 GlnGlyGlyGlnAspIleLeuSerLeuMetLysGlnMetLysProAr 267
|||||
763 CAAGTGGCCAAAGATATTTTGTCCCTTATGGCCAGATGATGAACACAG 812
267 gLysThrGluIleThrGluLysLeuArgGlnGluIleAsnLysValAla 284
|||||
813 AAAGACTGAATTCACCCGAAAACTACGCGCAAGAAATTAATATAGGTGTA 862
284 snArgTyrIleAspGluGlyIleAlaGluLeuValProGlyValIleuPhe 300
|||||
863 ATAGATATATCGATCAAGCAATGCGAGCTTGATACCTGCTGCTCAAAATAGG 912
301 lLeAspGluValHisMetLeuAspIleGluCysPheSerTyrLeuAsnAr 317
|||||
913 ATTGATGAGGTCCACATGTTGATATCGAATGTTTCTTATATCTTAACCG 962
317 gAlaLeuGluSerProLeuSerProIleValIleuAlaThrAsnArg 334
|||||
963 TGCAATGAGAGCCCATTAATATCCCAATGCTGATCTGCTACAAATAGG 1012
334 lYlLeCysAsnValArgGlyThrAspMetThrSerProHisGlyIlePro 350
|||||
1013 CAATATGTAATGTAAGAGCACTGATATGACAGTCCACATGATATACG 1062
351 ValAspLeuLeuAspArgLeuValIleIleArgThrGluThrTyrGlyPr 367
|||||
1063 GTGGATCTTCTAGATAGCGCTGTGATTAATTCGACAGACATATAGCCCC 1112
367 cThrGluMetIleGlnIleLeuAlaIleArgAlaGlnValGluGluIleA 384
|||||
1113 TACTGAGATGATACGATATGCTATCCGACACAAAGTGAAGAGATGG 1162
384 spMetAspGluGluSerLeuAlaTyrLeuGlyGlnIleGlyGlnIleThr 400
|||||
1163 ATATGATGAGAAAGAAAGTCTCTTATTAGCGAGATCCGACACACACA 1212
401 SerLeuArgHisAlaIleGlnLeuIleSerProAlaSerValValSerLy 417
|||||
1213 TCTTTGAGACATGCTATTCATTCATATCATCTCCAGCGGTGCTCAAA 1262
417 sThrAsnGlyArgGluLysIleCysLysAlaAspLeuGluValSerG 434
|||||
1263 GACTAATGGAAGAGAGAAATCTCAAAGCTGATCTCGAGGAAGTATGCTG 1312
434 lYleuTyrLeuAspAlaLysSerSerAlaArgLeuLeuGlnGlnGln 450
|||||
1313 GCGCTATTTTGATGCCAAATCTCGGCTCGGCTCGCTCCAGGAGACACAA 1362
451 GluArgTyrIleThr 455
|||||
1363 GAAAGATATCATTCACC 1377
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC38992
seq_documentation_block:
ID AAC38992 standard; DNA; 1612 BP.
XX
XX AAC38992;
AC
XX
DT 17-OCT-2000 (first entry)
DE
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 22954.
XX
XX Hybridisation assay: genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway: promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
XX PN EP1033405-A2.

```

```
XX 06-SEP-2000.
FD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132407.
PR 06-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0133256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144125.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145376.
PR 26-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147916.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
```

```
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159325.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment_scores:
  Quality: 1994.00      Length: 455
  Ratio: 4.542          Gaps: 0
  Percent Similarity: 96.484      Percent Identity: 85.055
```

```
alignment_block:
us-09-589-510-4 x AAC38992 ..
Align seg 1/1 to: AAC38992 from: 1 to: 1612
```

```
1 MetatglliegluGluValGlnSerThrseryslysglnatrgllleaIatH 17
  :::::::::::::::::::::
133 GTAAGATTGAAGAAATTCAGTCCACCGCTAAGAAACAAACGATTGCTAC 182
17 HlshrhshlslslslysglyleuGlyleuAspAlaasnlymetalailea 34
  :::::::::::::::::::::
193 TCACACCCATATCAAAAGCCTTGCGCTGACCAACTGTATCCCTATTA 232
34 laleuAlalaglypheValGlyGlnAlaAlaAlaIaIaIaIaIaIaIaIa 50
  :::::::::::::::::::::
233 AATTGGCAGCTGATTTGTTGTCACAACTTGAGGCTAGAGAGCAGCTGGT 282
51 leuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 67
  :::::::::::::::::::::
283 CTGTGAGTTGACATGATTAAACAGAAAGAAATGGCGGCAAGGCTCTTTT 332
67 uleuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 84
  :::::::::::::::::::::
333 GCTTCTGAGACTCTCTGGAAGCTGGGAAACAGCTTTGCTTGTGAATCT 382
84 laIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 100
  :::::::::::::::::::::
383 CTCAAGAGCTGGGAAGCAAGCTTCATCTGTCCAATGTGTGATCTGAG 432
```

```
101 ValTyrSerSerGluValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 117
  :::::::::::::::::::::
433 GTTACTCATCAGAGGTTAAGAAACAGAGCTTCTATGAGCAATTTTAg 482
117 gArGAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 134
  :::::::::::::::::::::
483 ACGTGCCATTGGTCTACGTATCAAGGAACCAAGAAAGTGTATGAAGGGG 532
134 luValThrGluLeuSerProGluGluAlaIaIaIaIaIaIaIaIaIaIaIa 150
  :::::::::::::::::::::
533 AGGACACGAGCTGTCCACCGAAGAAACTGAAGCCTCACTGAGGTTAT 582
151 AlAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 167
  :::::::::::::::::::::
583 GGTAAAGACATCAGCAGCTGTTGAATTACCTCAAGACATCAAGAAAGAAC 632
167 rLySGlnLeuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 184
  :::::::::::::::::::::
633 CAACATCTCAAAATTTGATCCCACTATCTATGATGCCCTTGATTAAGAAA 682
184 ysValAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 200
  :::::::::::::::::::::
683 AGGTAGCTGTAGAGAGATGTAATCTATATCGAAGCAAAACATGAGCCTGTC 732
201 LysArgValGlyArgCysAspSerPheAlaIaIaIaIaIaIaIaIaIaIaIa 217
  :::::::::::::::::::::
733 AAAGGGTAGTAGAGATGATGATCTTTGGCCCTCAATTTGATGTGAAGAG 782
217 agluGluTyrValProIleProIleProIleGlyGluValIaIaIaIaIaIaIa 234
  :::::::::::::::::::::
783 AGAAGATATGTTCCACTTCCCAAGAGAGAGTCCCAAAAGAAAGAGAGAG 832
234 leValGlnAspValThrLeuHisAspLeuAspAlaIaIaIaIaIaIaIaIaIa 250
  :::::::::::::::::::::
833 TAGTCGAGGATGTACACTCCAAAGATCTGGATCGACAAATGCTCCAGCCT 882
251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLeuProIle 267
  :::::::::::::::::::::
883 CAAGGTGGCCAGGATATCTTCTTTGATGGGCCCAATGATGAACCGCG 932
267 glyThrGluIleThrGluIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 284
  :::::::::::::::::::::
933 GAAGACGAGATCAGTATAGCTTCGGCAAGAAATTAACAAAGCTGTGA 982
284 snArgTyrIleAspGluGlyIleAlaIaIaIaIaIaIaIaIaIaIaIaIaIa 300
  :::::::::::::::::::::
983 ACCGATATATAGATGAAGGTGTGGCAGACGCTGTCCAGAGCTTCTATTT 1032
301 IleAspGluValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 317
  :::::::::::::::::::::
1033 ATTGATGAGGTTCAATATGCTTGAATGAGTGTCTTCATCTGTAACCG 1082
317 gaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 334
  :::::::::::::::::::::
1083 TGCTCTTGAGAGCCATATATCTCCGATAGATATTTGCCAACAATAGAG 1132
334 lyIleCysAsnValArgGlyThrAspMetThrSerProHisGlyIlePro 350
  :::::::::::::::::::::
1133 GTGTTTGCACAGTAAAGGAGCATATGCCCCAGCCCCCTGAGAGTCCCT 1182
351 ValAspLeuLeuAspArgLeuValIleIleArgThrGluThrTyrGlyPyr 367
  :::::::::::::::::::::
1183 ATTGATATATATATGATCATGCTGTTATCATCCGACCTCAAAATCTATATCC 1232
367 oThrGluMetIleGlnIleLeuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 384
  :::::::::::::::::::::
1233 CTCTGAATATGATCCAGATATATAGCAATTCGTGGCAAGTTGAAGAAATTA 1282
384 spMetAspGluGluSerLeuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 400
  :::::::::::::::::::::
1283 CCGTGTATGAAGAAATCTGGTCTACTTGGGAGATGTGGCCAAAGAAAGCT 1332
```



```

317  gAlaLeuGIuSerProLeuSerProIleValIleLeuAlaThrAsnArg 334
      |||.....|
1010  CGCCCTGAGCTTCTTATCGCTCCATCGATCTTGCATCCAGACGAG 1059
      |||.....|
334  LyIecySAsnValArgIlyThr...AspMetThrSerProHisGlyIle 349
      |||.....|
1060  GCAACTGTGTCAGAGCAGTCAGATCATCCCTCAGCGGATC 1109
      |||.....|
350  ProValAspLeuAspArgLeuValIleIleArgThrGluThrTyrgI 366
      |||.....|
1110  CCTCTTACCTTCTGACCGAGTGATCATCCGACCATGCTGTATAC 1159
      |||.....|
366  yProThrGluMetIleGluIleLeuAlaIleArgAlaGluValGluGlu 383
      |||.....|
1160  TCCACAGGAATGAAACAGATCATTAATCCCTGCCACAGAGAGAA 1209
      |||.....|
383  IeAspMetAspGluGluSerLeuAlaIleArgIleGluIleGlyGln 399
      |||.....|
1210  TCAACATCAGTGAAGAGCAGCTGACACCTGGGGAGATTGGACCAAG 1259
      |||.....|
400  ThrSerLeuArgHisAlaIleGluIleSerProAlaSerValAlaSe 416
      |||.....|
1260  ACCACACTGAGGTACTGAGTGCAGCTGACCCCGCCCAACTGCTTGC 1309
      |||.....|
416  rLySThrAsnGlyArgGluIlySlyIleCysAlaAspLeuGluValS 433
      |||.....|
1310  TAAATATCAGCGGAGAGACAGCATTTGAGAAAGACATGTGAAAGATCA 1359
      |||.....|
433  erGlyLeuTyrlLeuAspAlaIlySerSerAlaArgLeuGluGln 449
      |||.....|
1360  TGAACCTTTCTATGATGCCAGTCCCTCCGCAAAATCCTGGGTGACAG 1409
      |||.....|
450  GlnGluArgTyrlle 454
      |||.....|
1410  CAGGATTAAGTACATG 1424

```

```

seq_name: /sids1/gcgdata/geneseq/geneseqn/NAI999.DAT:AAx18198
seq_documentation_block:
ID   AAX18198 standard; cDNA to mRNA; 1587 BP.
AC   AAX18198;

```

```

XX  10-MAY-1999 (first entry)
XX  TIP49 coding sequence.
XX  TIP49; TATA binding protein; TBP binding protein;
XX  DNA transcription control; ss.
XX  Rattus sp.
XX  WO9900419-A1.
XX  07-JAN-1999.
XX  25-JUN-1998; 98WO-JP02836.
XX  27-JUN-1997; 97JP-0187398.
XX  (SUME ) SUMITOMO ELECTRIC IND CO.
XX  Kishimoto T, Makino Y, Niwa S, Tamura T;
XX  WPI; 1999-095682/08.
XX  P-PSDB; AAW74416.
XX  TBP-binding protein with DNA helicase and ATPase activities - gene
XX  encoding it, and antibodies recognising it.
XX  Claim 10; Page 39-41; 64pp; Japanese.

```

CC This sequence encodes the TATA-binding protein (TBP) binding protein,
 CC designated TIP49, of the invention. TIP49 and its associated nucleic
 CC acids and antibodies are useful in investigation of the process of DNA
 CC transcription control by TBP in vivo, and in detection of the blocking of
 CC DNA transcription.
 XX
 SQ Sequence 1587 BP; 439 A; 362 C; 464 G; 322 T; 0 other;

alignment_scores:
 Quality: 1722.50 Length: 455
 Ratio: 4.11 Gaps: 1
 Percent Similarity: 92.088 Percent Identity: 72.308

alignment_block:
 US-09-589-510-4 x AAX18198 ..

Align seg 1/1 to: AAX18198 from: 1 to: 1587

```

1 MetArgIleGluGluValGlnSerThrSerLysGlnArgIleAlaThr 17
  |||.....|
31 ATGAGATTGAGGAGGTGAGAGCACCGAAGACGACGACGATGCTC 80
  |||.....|
17 rHisThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleA 34
  |||.....|
81 CCACAGCCACGCTGAAGGGCTGGGCTGATGAGAGCGGCTGGCCAAAGC 130
  |||.....|
34 IeLeuAlaIleGlyPheValGlyGlnAlaAlaIleArgGluAlaIleGly 50
  |||.....|
131 AGCGGCTTCGGGGCTCGTGGCCAGAGAACGAGAGAGGACATGTGT 180
  |||.....|
51 LeuAlaIleAspMetIleArgIleLysLysMetAlaIleArgAlaValle 67
  |||.....|
181 GTCATAGTGGAAATTATCTCAAAAGCAAGAAATGCTGGAAAGCTGTCT 230
  |||.....|
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
  |||.....|
231 GTTGGCAGGGCTCCTCGAAGCTGGCAAGACAGCTTGGCCCTGGCTATTG 280
  |||.....|
84 IeGlnGluLeuGlySerLysValPropheCysProMetValGlySerGlu 100
  |||.....|
281 CTCAGGAACCTGGGCGAGTAAAGTCCCTTCTGCCGATGTGGGTAGTGA 330
  |||.....|
101 ValTyrlSerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
  |||.....|
331 GPTACTCACTGAGATCAGAGACAGAGGCTGTGATGGAACCTTCG 380
  |||.....|
117 gArgAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrlGlyG 134
  |||.....|
381 AAGGCTATTGGGCTGGCGATTAAGAGAGCTTAAGAGGTTATGAAGGG 430
  |||.....|
134 IuValThrGluLeuSerProGluIleAlaGluSerThrGlyGlyTyrl 150
  |||.....|
431 AGGTACAGAGCTCACTCCCTGTGAGACAGAGAACCCCATGGGTGTAT 480
  |||.....|
151 AlaLysSerIleSerHisValIleIleSerLeuLysThrValGlyGlyTh 167
  |||.....|
481 GGCAGAAACTATCAGCCACGTGATCATAGGGCTCAAGACTGCCAAAGAGAC 530
  |||.....|
167 rLySglnLeuLysLeuAspSerSerIleTyrlAspAlaLeuIleLysGlu 184
  |||.....|
531 CAACACAGCTGAAGCTGAGCCCGCATATTTTTGAAGTTTGCAGAAAGAAC 580
  |||.....|
184 ySValAlaValGlyAspValIleTyrlIleGluAlaAsnSerGlyAlaVal 200
  |||.....|
581 GAGTAGAGGCTGAGAGATGATTTACATTGAAGCAAAACAGTGAAGCTGTG 630
  |||.....|
201 LysArgValGlyArgCysAspSerPheAlaThrGluTyrlAspLeuGluAl 217
  |||.....|
631 AAGAGGCAAGGAGGTGTGACACCTATGCCACAGAGCTTGAAGCTTGAAGC 680
  |||.....|
217 aGluIuTyrlValProIleProLysGlyGluValHisLysLysGluI 234
  |||.....|

```

```

681 TGAAGATATGTCCTTGGCAAGGAGATGTGCAAGAAAGAGAGAA 730
234 lvalglinspvalthrleuHisaspLeuaspAlaIalaasnaglinspro 250
731 TCATACAGGATGTGACCTTGCACGACTTGACCTGGCAATGGCGGCTT 780
251 GlnGlyGlnGlnAspIleLeuSerLeuMetGlyGlnMetLeuProAr 267
781 CAGGTGGGCAAGATATTTCTGTATGATGGCCAGTTGATGAAGCCAA 830
267 glyThrGlnIleThrGlnLysLeuArgGlnGlnIleAsnLysValVala 284
831 AAAGACAGAGATCACAGATTAACCTGACGGGAGATCAACAAAGTGTTGA 880
284 snArGTYrIleAspGlnGlyIleAlaGlnLeuValProGlyValLeuPhe 300
881 ACAAAATRCATGTGACCAAGGGCTTGCAAGCTGTCCCTGGAGTGCCTTT 930
301 lIeAspGlnValHisMetLeuAspIleGluCysPheSerTYrLeuAsnAr 317
931 GTTGACGAGGTCCACATGCTGATATCGATGCTTACCTACCTGCACCG 980
317 gaIaLeuGlnUserProLeuSerProIleValIleLeuAlaIthrAsnArg 334
981 AGCCCTGGAGTCCCATCGCCCAATTCATCTTTCATCTTTCATCAACCGAG 1030
334 lylIecYasnValArgGlyThr...AspMetThrSerProHisGlyIle 349
1031 GCACAGTGTGCATCAGGCGCAGCGAGACATCATCTTCCACACGGCATC 1080
350 ProValAspLeuLeuAspArgLeuValIleIleArgThrGluThrTyrG1 366
1081 CCTTGTGACCTGCTGACCGGGGATGATCATCAGACCATGCTGTATAC 1130
366 yProThrGlnMetIleGlnIleLeuValIleArgGlnIleValGlnGlnI 383
1131 GCCACAGGAGATGACAGATCATTAAGATCCAGGCCAGCGAGAGGCA 1180
383 lEaspMetAspGlnGlnUserLeuAlaIthrLeuGlyGlnIleGlnGln 399
1181 TCAACATCATGAGTGAAGGCGCTTAACACCTCGGGAGATGCGACCAAG 1230
400 ThrSerLeuArgHisAlaIleGlnLeuIleSerProAlaSerValValSe 416
1231 ACCACCTGTGATTCAGTGCAGCTGTGACCCCTGCCAACCTGCTGTGC 1280
416 rlyStrAsnGlyArgGlnLysIleCysLysAlaAspLeuGlnGlnValS 433
1281 CAAGATCAACGGGAGAGACACGCTTGAAGAAGGACGCTGGAGAGATCA 1330
433 ergIleuTYrLeuAspAlaLysSerSerAlaArgLeuLeuGlnGln 449
1331 GCGAGCTCTTCTATGACGCCAAGTCTCCGCCAAGATCTGGGCGACGAG 1380
450 GlnGlnArgTYrIle 454
1381 CAGGACAAAGTACATG 1395
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAx24059

```

```

seq_documentation_block:
ID AAX24059 standard; DNA; 1750 BP.

```

```

AC AAX24059;
XX
XX 30-JUN-1999 (first entry)
DE Human regulatory molecule HRM-1 DNA.
XX
XX Human regulatory molecule; HRM-1; cytosolic activity; immune modulator;
KW transcription factor; enhancer; cell proliferation stimulation; cancer;
KW treatment; microarray; detection; diagnosis; cell proliferation disease;
KW leukemia; lymphoma; myeloma; adenocarcinoma; sarcoma; bladder; bone;

```

```

KW brain; lung; liver; ovary; skin; teratocarcinoma; immune response;
KW allergy; asthma; diabetes; multiple sclerosis; Grave's disease;
KW myasthenia gravis; ss.
OS Homo sapiens.
PN WD9915658-A2.
XX
XX 01-APR-1999.
PD
XX 22-SEP-1998; 98WO-US19839.
XX
XX 23-SEP-1997; 97US-0933750.
XX
PA (INCY-) INCYTE PHARM INC.
PI Au-Young J, Bandman O, Guegler KJ, Hillman JL, Lal P;
PI Shah P, Yue H;
XX
XX WPI; 1999-254710/21.
DR P-PsDB; AAW93945.
XX
PT New human regulatory molecules
PS
PS Claim 5; Page 70-71; 76pp; English.
XX
CC This invention describes novel human regulatory molecules (HRM) which
CC have cytosolic activity and act as immune modulators, transcription
CC factors or enhancers. The HRMs can be used to stimulate cell
CC proliferation. Antagonists and agonists of the proteins of the invention
CC can be used to treat cancer. The encoding nucleic acids can be used in
CC microarrays to detect polynucleotides (and their expression levels) that
CC encode HRMs in a biological sample. The HRMs and microarrays can be used
CC to diagnose, treat or prevent cell proliferation diseases especially cancer,
CC e.g. leukemia, lymphoma, myeloma, adenocarcinoma, sarcoma, cancer of e.g.
CC bladder, bone, brain, lung, liver, ovary, skin, etc, teratocarcinoma, or
CC to treat or prevent immune responses e.g. allergies, asthma, diabetes,
CC multiple sclerosis, Grave's disease or myasthenia gravis.
XX
SQ Sequence 1750 BP; 472 A; 406 C; 485 G; 387 T; 0 other;

```

```

alignment_scores:
Quality: 1721.50 Length: 455
Ratio: 4.109 Gaps: 1
Percent Similarity: 92.088 Percent Identity: 72.527

```

```

alignment_block:
US-09-589-510-4 x AAX24059 ..

```

```

Align seg 1/1 to: AAX24059 from: 1 to: 1750

```

```

1 MetArgIleGlnGlnValGlnSerThrSerLysLysGlnArgIleAlaThr 17
|||||
82 ATGAAGATTGAGAGGTGAAGACATCAAGACGACGACATGGCTTC 131
17 rHisThrHisIleLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 34
|||||
132 CCACAGCCACGCTGAAGAGGCTGGGCTGGAGACGAGCGGCTGGCCAAAGC 181
34 lAlaValAlaAlaGlyPheValAlaGlnAlaAlaAlaArgGlnAlaAlaGly 50
|||||
182 AGGCGGCTCAAGGCTGTGGGCTGGGCAAGACGCGCTGGAGGATGTGGC 231
51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlnArgAlaValLe 67
...
232 GTCATAGTAGAATTAAATCAAGCAAGAAATGGCTGGAAGAGCTGTCTT 281
67 lLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
|||||
282 GTTGCGAGGACCTCTGGAAGTGGCAAGACGCTGTGGCTGTATTTG 331
84 lAlaGlnLysLysSerLysValProPheCysProMetValGlySerGln 100

```

```

|||||
332 CTCAGAGAGCTGGTAGTAGAGTCCCTCTGCGCAATGCTGGGAGAGTGA 381
101 VALTYRSErSerGluValLysLysThrGluValLeuMetGluAsnPro 117
382 GTTTCCTCACTGAGATCAAGAGACAGAGGCTGATGAGAGACTTCCG 431
117 gATgAlaIleGlyLeuArgLysGluAsnLysGluValTyrGluGly 134
432 CAGGCGCATTTGGGCTGCGAATTAAGAGACCAAGCAAGTTATGAGAG 481
134 LuValThrGluLeuSerProGluGluAlaGluSerThrThyGlyGly 150
482 AAGTCACAGAGCTAACTCCGTGTGAGACAGAGCAATCCATGGAGAT 531
151 AlaLysSerLysSerHisValIleLeuSerLeuLysThrValLysGly 167
532 GGCAGAACCTTAGCCATGTGATCATAGCACTCAAAACAGCCAAAGAAC 581
167 rLysGluLeuLysLeuAspSerSerLLeTyrAspAlaLeuIleLys 184
582 CAACAGAGTTGAACACTGGACCCAGCATTTTGAAGTTTGCAGAAAGAG 631
184 yValAlaValAlaGlyAspValIleTyrIleGluAlaAsnSerGlyAla 200
632 GAGTAGAAGCTGGAGATGATTTACATTGAGCCAGCAACCTGGGCGCTG 681
201 LysArgValAlaGlyArgCysAspSerPheAlaThrGluTyrAspLeu 217
682 AAGAGCGAGGCGAGTGTGATCTACTGTCACAGAAATTCGACTTGAAG 731
217 agLugLutYrValProIleProLysGlyLysValHisLysLysGly 234
732 TGAAGATGATGCTCCCTGCCAAAGGAGATGCGCAAAAGAAAGAA 781
234 LeValGluAspValThrLeuHisAspLeuAspAlaAlaAsnLagIn 250
782 TCATCCAGAGATGACCTTCCATGACTGATGATGCTTAATCGCGGCC 831
251 GluGlyGlyGluAspIleLeuSerLeuMetGlyGluMetLysPro 267
832 CAGGGGAGCAAGATATCCGTCCATGATGGCGCACCTAATGAGCCAA 881
267 gLysThrGluIleThrGluLysLeuArgGluGluIleAsnLysValAla 284
882 GAAGACAGAAATCACAGCAAACTCGAGGGGAGATTAAAGTGCTGA 931
284 snArgTyrIleAspGluGlyIleAlaGluLeuValProGlyValLeu 300
932 ACAAGTACATCGACCGAGCATTCGTAGCTGCTCCGGGTGCTGTT 981
301 IleAspGluValHisMetLeuAspIleGlyCysPheSerTyrLeuAsn 317
982 GTTGATGAGGTCACATGCTGAGCATGATGATGCTTCACTACCTGCA 1031
317 gaLaLeuGluSerProLeuSerProIleValIleLeuAlaThrAsnArg 334
1032 CGCCCTGAGCTTCTATCGCTCCATCGTATCTTGGCAATCCACCGAG 1081
334 LyIleCysAsnValArgGlyThr...AspMetHisSerProHisGly 349
1082 GCAACTGTGATCATGAGAGCGACTGAGACATCAATCCCTCAGCGGAT 1131
350 ProValAspLeuLeuAspArgLeuValIleLeuArgThrGluThrArg 366
1132 CCTTTCAGCTTCTGAGCGAGTATATATCGGACCACTGCTGTATAC 1181
366 yProThrGluMetIleGluIleLeuAlaIleArgAlaGluValGluGlu 383
1182 TCACAGAGAAATGAACAGATCATTAATAATCCGCGCCAGACGAGAGAA 1231
383 LeAspMetAspGluGluSerLeuAlaTyrLeuGlyGluIleGlyGlu 399
|||||

```

```

1232 TCACATCATGATGAGGAGCGACTGAACACACTGGGGGAGATTGGACCCAG 1281
400 ThrSerLeuArgHisAlaIleGluLeuIleSerProAlaSerValIse 416
1282 ACCACACTGAGTACTCTCAGCTGAGCTGCTGACCCCGGCAACTTGTG 1331
416 rLysThrAsnGlyArgGluLysIleCysLysAlaAspLeuGluLys 433
1332 TAAATCAACGGGAGGAGCAGCATTTGACAAAGAGCATGTGCAAGAGATCA 1381
433 erGlyLeuTyrLeuAspAlaLysSerSerAlaArgLeuLeuGluGlu 449
1382 GTGAATTTTCTATGATGCCAAGTCTCCGCCAAATCTGCTGACACAG 1431
450 GlnGluArgTyrIle 454
1432 CAGGATACGTACATG 1446

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC59220
seq_documentation_block:
ID AAC59220 standard; CDNA: 1744 BP.
XX
XX AAC59220;
XX
XX 29-JAN-2001 (first entry)
XX
XX Human secreted protein cDNA sequence #6.
XX
XX Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; ss.
XX
XX Homo sapiens.
XX
XX WO20005199-A1.
XX
XX 21-SEP-2000.
XX
XX 09-MAR-2000; 2000MO-US06014.
XX
XX 12-MAR-1999; 99US-0124095.
XX 11-JUN-1999; 99US-0138598.
XX 03-DEC-1999; 99US-0168665.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI: 2000-572359/53.
XX P-PSDB; AAB27799.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition
XX
XX Claim 1; Page 343; 433pp; English.
XX
XX The invention relate to the isolation of genes AAC59215-C59261 encoding
XX 47 human secreted proteins AAB27794-B27840. The genes can be used to
XX generate fusion proteins by linking to the gene for the human
XX immunoglobulin G Fc portion for increasing the stability of
XX the fusion protein as compared to the human protein only. The genes and
XX proteins are useful for preventing, ameliorating or treating medical
XX conditions, e.g. by protein or gene therapy. The genes are isolated
XX from a range of human tissues disclosed in the specification. The
XX nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

```

CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischemias; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.

XX Sequence 1744 BP; 476 A; 403 C; 480 G; 384 T; 1 other;

alignment_scores:

Quality: 1570.00 Length: 458
 Ratio: 3.811 Gaps: 3
 Percent Similarity: 89.956 Percent Identity: 70.306

alignment_block:

US-09-589-510-4 x AAC59220 ..

Align seg 1/1 to: AAC59220 from: 1 to: 1744

```

1 MetArgIleGluGluValGlnSerThrSerLysGlnArgIleAlaThr 17
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 ATGAGAGATTGAGGAGTGAAGACACTACGACAGCGCATCGCTC 125
17 rHsThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaLea 34
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
126 CACAGCGCAGGTGAAGAGGCTGGGCTGACAGAGCGGCTGGCCAAGC 175
34 LalaValAlaGlyPheValGlyGlnAlaAlaArgGluAlaIleGly 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
176 AGGAGGCTCAGGGCTGTGGGCGCAGAGAACCGCGAGAGCATGTGGC 225
51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaVal 67
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
226 GTCATGATGAATTAATCAAAAGCAAAATGCTGGAAGAGCTGTCTT 275
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIle 84
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
276 GTTGGCAGAGACCTCCTCGAAGCTGGCAAGACAGCTGTGGCTATTG 325
84 IalGlnGluLeuGlySerLysValProPheCysProMetValGlySerGlu 100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
326 CTCAGAGAGCTGGGTAGTAAAGTCCCTCTGCCAATGGTGGGAGTGAA 375
101 ValTyrSerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
376 GTTACTCTACCTGAGATCAAGAAAGACAGAGGTGATGGAAGACTTCG 425
117 gArgAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGluGly 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
426 CAGGCC.ATTGGGCTGCGAATAAAGAGACCAAGAGATTATTGAGT. 473
134 LuValThrGluLeuSerProGluGluAlaGluSerThrThrGlyTyr 150
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
474 AAGTCAACAGAGCTAACTCCGTGTGAGACAGAGAAATCCCATGSGAGATAT 523
151 AlaLysSerIleSerHisValIleIleSerLeuLysThrValLysGlyTh 167
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
524 GCGAAACCATTAAGCCATGTGATCATAGACTCAAAACAGCCAAAGGAGAC 573
167 rLysGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGlu 184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
574 CAACAGTTGAAGACTGAGCCAGCATTTTGAAGTTTGCAGAAAGAGC 623
184 ysValAlaValGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
624 GAGTGAAGAGCTGAGATGTGATTACATTGAAGCCAAACAGTGGGCGGTG 673
201 LysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGluAl 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
674 AAGAGCAGAGGCGAGGTGATCACTATGCCACAAATTCGACTTGAAGC 723
217 aGluGluTyrValProIleProLysGlyGluValHisLysLysGluI 234
   |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

724 TGAAGATATGTCCTTGGCCAAAAGGGATGTGCACAAAAGAAAGAAA 773
234 LeValGlnSerValThrLeuHisAspLeuAspAlaAlaAsnIleGlnPro 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
774 TCATCCAGATGTGACCTTGACATGACTTGATGTGGCTAATGCGCGGCC 823
251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLysProAr 267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
824 AG.GGGGACAAAGATATCCTGTCCATGATGGGCGACGTAATGAAGCCAA 872
267 GlyThrGluIleThrGluLysLeuArgGln.....GluIleAsnL 281
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
873 AAGA.....CAGAAATCAGACAAACTTCAGGGAGATTAATA 910
281 ysValValAsnArgTyrIleAspGluGlyIleAlaGluLeuValProGly 297
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
911 AGTGTGTGAACAAGTACATCGACAGGCGATGTGAGCTGTGCTCCGGGT 960
298 ValLeuPheIleAspGluValHisMetLeuAspIleGluCysPheSerTy 314
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
961 GTGCTGTTGTGATGAGGTCCATGTCTGACATTTGAGTCTTCACCTA 1010
314 rLeuAsnArgAlaLeuGluSerProLeuSerProIleValIleLeuAla 331
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1011 CTGCAACGCGCGCTGAGTCTTCTATGCTGCCATCTCATCTTTGCAT 1060
331 hAsnArgGlyIleCysAsnValArgGlyThr...AspMetThrSerPro 346
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1061 CCACACGAGGCAACTGTGTCTATCAAGGACCTGAGACATCACATCCCT 1110
347 HisGlyIleProValAspLeuAspArgLeuValIleIleArgThrGlu 363
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1111 CACGGCATCCCTCTTGRCTTCTGACCGAGATGATTAATCCGGACCAT 1160
363 uThrTyrGlyProThrGluMetIleGlnIleLeuAlaIleArgAlaGln 380
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1161 GCTGATATCTCCACAGAAATGAATCAATCATTAATTCGTCGCCCA 1210
380 aGluGluIleAspMetAspGluGluSerLeuAlaTyrLeuGlyGluIle 396
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1211 CGGAAGG.ATCAACATCAAGTGAAGAGGACTGAACCACTGGGGGAGATT 1259
397 GlyGlnGlnThrSerLeuArgHisAlaIleGlnLeuLeuSerProAla 413
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1260 GGCACCAAGACCACTGAGGTACTCAGTGCAGTGTGACCCCGGCCAA 1309
413 rValValSerLysThrAsnGlyArgGluLysIleCysLysAlaAspLeu 430
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1310 CTTGCTTGCTAAATCAACGGGAGACACAGCATTTGAAGAAAGCATGTCG 1359
430 LuGluValSerGlyLeuTyrLeuAspAlaLysSerSerAlaArgLeuLeu 446
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1360 AAGAGATCAAGAGACTTTTCTATGATGCCAAGTCTCCGCAAAATCTCG 1409
447 GlnGluGlnGlnGluArgTyrIle 454
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1410 GCTGACCAAGCAGATAGTACATG 1433

```

seq_name: /stids/gcgdata/geneseq/geneseqn/MA1999.DAT:AAZ52915

seq_documentation_block:

ID AAZ52915 standard; cDNA, 1509 BP.

XX AAZ52915;

DT 14-MAR-2000 (first entry)

DE Human prostate tumor cDNA library derived EST fragment #58.

KW Pancreas; tumor; EST; expressed sequence tag; human; cytosolic;

XX treatment; ds.

OS Homo sapiens.

XX AAF13700;
 AC
 XX 13-MAR-2001 (first entry)
 DT
 XX Aspergillus oryzae EST SEQ ID NO:6223.
 DE
 XX Multiple gene expression; filamentous fungal cell; EST;
 XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey WM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
 DR WPI, 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT using fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 88; Page 2561; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organization of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus
 CC AAF14879 to AAF15337 represents ESTs from Aspergillus oryzae; and
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 1213 BP; 336 A; 325 C; 275 G; 276 T; 1 other;

alignment_scores:

Quality: 1064.50 Length: 297
 Ratio: 3.972 Gaps: 1
 Percent Similarity: 90.236 Percent Identity: 68.687

alignment_block:

us-09-589-510-4 x AAF13700 ..
 Align seg 1/1 to: AAF13700 from: 1 to: 1213

117 ArgArGAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGluGlu 133
 1 CGAGAGCGCATGTGGTCTCCGTCGCGTAACCAACCAAGAGTACCAAG 50
 133 yGluValThrGluLeuSerProGluGluAlaGluSerThrThrGlyGlyT 150
 51 CGAAGTCACCGACCTTAACGCTGAGCAAACTGAAATCCATTGGGAGGCT 100
 150 yAlaLysSerIleSerHisValIleIleSerLeuLysThrValLysGly 166
 101 ATGACCCCAACATACGACCTTATTTGATGATGAGCTGCAAAAGCA 150
 167 ThrLysGluLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysG 183
 151 ACCAAGACGTCGCGCTGATCCAGCATTTTGAAGGCTATTCAAAAGA 200
 183 uLysValAlaValGlyAspValIleTyrIleGluAlaAsnSerGlyAlaV 200
 201 ACCAGTCACTGTTGGAGATGTCATCTACATCGAACCAACTGGTGCTT 250
 200 aLysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGlu 216
 251 GTAAGAGACTGGCGCGATGCCGATGCCGTATCGACGAGATTGACCTTAA 300
 217 AlaGluGluTyrValProIleProLysGluValHisLysLysLysG 233
 301 GCAGAGAAATACGTCCTGTACCAAGGAGGTTCCACAAGAAAGA 350
 233 uLleValGlnAspValThrLeuHisAspLeuAspAlaLysnLaglnP 250
 351 AATCGTAGCAGGACGTGACGCTACATGACCTTGACATGGCCACCTCGGC 400
 250 rGlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLysPro 266
 401 CACAAGTGGACAGGACGCTATGACCATGATGAGCAACATGATGAACCC 450
 267 ArgLysThrGluIleThrGluLysLeuArgGlnLysLysValVa 283
 451 AAGAAGACCGAGATCCGATGACCTACGTCAGAGATGACACAAGGTTGT 500
 283 lAsnArgTyrIleAspGluGlyIleAlaGluLeuValProGlyValLeuP 300
 501 TAGCGGTATATCGACCAAGAGTGGTGAACCTGTTCTCGGTCTCT 550
 300 heLleAspGluValHisMetLeuAspIleGluCysPheSerTyrLeuAsn 316
 551 WTATTGATGAGGTCCACATGTTGGATTGAAATGTTTCAACCTATCTCAC 600
 317 ArgAlaLeuGluSerProLeuSerProIleValIleLeuAlaThrAsnR 333
 601 CGGCGACTTGAACTCTCAATCTCCCATTTGATATCTCCCTCCACCG 650
 333 gGlyTLeCysAsnValArgGlyThr...AspMetThrSerProHisGlyT 349
 651 CGGCCACACTGTGATCCGCGGACCGACGACATCAAGCGCGGACAGGA 700
 349 leProValAspLeuLeuAspArgLeuValIleIleArgThrGluThrTyr 365
 701 TTCCCTGTGACCTCTGCTGCTGCTCATCATCCACCCACACCTTAC 750
 366 GlYProThrGluMetIleGlnIleLeuAlaIleArgAlaGluValGluG 382
 751 TCTCCGAGAGATCAAGACCATCTTCGCTTCGCGCCAAACAGAAAG 800
 382 uLleAspMetAspGluGluSerLeuAlaTyrLeuGluGlyLysGlyGln 399
 801 TCTCAACATTACCGATCCGCTCCGACAAGATCTCCGACGACGAGCA 850
 399 lThrSerLeuArgHisAlaIleGlnLeuIleSerProAla 412
 851 ACGTCAAGCTGCGGTACGACCTGACATTAAGTACCCAGCT 891

```

seq_name: /SIPSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA90316
seq_documentation_block:
ID   AAA90316 standard; cDNA; 1492 BP.
AC   AAA90316;
XX
XX   22-DEC-2000 (first entry)
DE   Human HEL50 DNA helicase cDNA.
XX
XX   Human; HEL50; DNA helicase; TIP49 homologue; EST AA374580;
KM   expressed sequence tag; antibody; antisense therapy; diagnosis;
KM   helicase-associated disease; genetic disorder; tumour; cancer; ss.
XX
XX   Homo sapiens.
XX
XX   Key   Location/Qualifiers
FH   CDS   1..1392
FT   /*tag= a
FT   /product= "Human HEL50"
FT   /function= "DNA helicase"
XX
XX   WO200047731-A1.
XX   17-AUG-2000.
XX
XX   22-NOV-1999; 99WO-JP06519.
XX
XX   10-FEB-1999; 99JP-0033052.
XX
XX   (SUME ) SUMITOMO ELECTRIC IND CO.
XX
XX   Tamura T;
XX
XX   WPI: 2000-543587/49.
XX   DR   P-PSDB; AAB21093.
XX
XX   DNA helicase of human origin, antibodies to it and DNA encoding it for
XX   investigation diagnosis and treatment of genetic disorders involving
XX   helicase gene expression
XX
XX   Claim 6; Page 41; 49p; Japanese.
XX
XX   This sequence represents cDNA encoding a novel human DNA helicase,
XX   HEL50, which has homology with a part of human TIP49. cDNA encoding
XX   human HEL50 was initially identified in an EST (expressed sequence tag)
XX   database as EST AA374580. The invention relates to human and yeast HEL50
XX   cDNA helicases (AAB21093, AAB21094), and to cDNAs encoding them
XX   (AAA90316, AAA90317). It also encompasses HEL50 antisense
XX   oligonucleotides and anti-HEL50 antibodies. HEL50 nucleotides, proteins
XX   and antibodies may be used for the investigation, diagnosis and treatment
XX   of diseases with which DNA helicase is associated, such as genetic
XX   disorders involving variant forms of the gene, and tumours in which
XX   altered levels of expression of the DNA helicase gene occur.
XX
XX   Sequence 1492 BP; 352 A; 439 C; 454 G; 247 T; 0 other;

alignment_scores:
Quality: 973.00 Length: 455
Ratio: 2.845 Gaps: 4
Percent Similarity: 75.165 Percent Identity: 43.956

alignment_block:
US-09-589-510-4 x AAA90316 ..
Align seq 1/1 to: AAA90316 from: 1 to: 1492
2 ArgIleGluGluValGlnSerThrSerLysGlnArgIleAlaThrH 18
::: ||| | | | | | | | | | | | | | | | | | | | | |
25 AAAGTCCCGGAGATCGGTGATGTACACAGGATTGAGCGAATCGTGCCCA 74

```

```

18 sThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleAla 35
::: ||| | | | | | | | | | | | | | | | | | | | | |
75 CTCACACATCCGGGAGCTGGGCTGGAGCATGCTTGGAGCCTCGGCAGG 124
35 euAlaIaGlyPheValIleGlyAlaIleAlaIaArgIleValIaGlyLeu 51
::: ||| | | | | | | | | | | | | | | | | | | | | |
125 CTCGCAAGGAGCATGTGGTGTACCTGGCGGCACGGCGGCGCTGGCGT 174
52 AlaValAspMetIleArgIleLysLysMetAlaGlyArgAlaValLeuLeu 68
::: ||| | | | | | | | | | | | | | | | | | | | | |
175 GTGCTGGAGATGATCCGGGAAGAGATTCGCGTGGCGAGCTCTTAT 224
68 uAlaGlyProProAlaThrGlyLysThrAlaAlaLeuGlyIleAlaG 85
::: ||| | | | | | | | | | | | | | | | | | | | | |
225 TGTGTCGACCGCGGCGACGGGGAAGACGGCATTCGCTGATGGCATGGCGC 274
85 IngIleuGlySerLysValProPheCysProMetValIleGlySerGlyVal 101
::: ||| | | | | | | | | | | | | | | | | | | | | |
275 AGGCCCTGGGCGCTGACACCGCATTCACAGCCATCGCGGCGAGTGAATC 324
102 TyrSerSerGlyValLysLysThrGluValLeuMetGluAsnPheArg 118
::: ||| | | | | | | | | | | | | | | | | | | | | |
325 TTCTCCCTGGAGATGACCAAGACCGAGCGCTGACGCGAGCTTCGCGCG 374
118 gAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGluGlyLeu 135
::: ||| | | | | | | | | | | | | | | | | | | | | |
375 GTCCATCGGCGCTGCATCAAGAGAGAGACGAGATCATGCAAGGGAGG 424
135 AlThrGluLeuSerProGluGluValLeuGlySerThrGlyGlyTyrAla 151
::: ||| | | | | | | | | | | | | | | | | | | | | |
425 TGTGGAGATCCAGATGATGACGACGACGAGGAGGCGGTCCAGAGTGT 474
152 LysSerIleSerHisValIleIleSerLeuLysThrValLysGlyThr 168
::: ||| | | | | | | | | | | | | | | | | | | | | |
475 GCGAAA.....CTGACCTCAAGACACAGCAGATGAGGAGAC 509
168 sGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGlyLys 185
::: ||| | | | | | | | | | | | | | | | | | | | | |
510 CATCTACGACCTGGGACCAAGATGATGATGCTCCCTGACCAAGCAAG 559
185 AlAlaValIleAspValIleTyrIleGluAlaAsnSerGlyAlaValLys 201
::: ||| | | | | | | | | | | | | | | | | | | | | |
560 TCAGAGCGCGGGAGCGTGTATCACCATCCACAGCGGAGCGGCAAGATCTCC 609
202 ArgValIleTyrGlyCysAspSerPheAlaThrGluTyrAspLeuGluAla 218
::: ||| | | | | | | | | | | | | | | | | | | | | |
610 AAGCTGGGCGCTCTCTTCACACGCGCCCGCAGCTACGAGCTATGGGCTC 659
218 uGlu.....TyrValProIleProLysGlyGluValHisLysLysLys 233
::: ||| | | | | | | | | | | | | | | | | | | | | |
660 CCAGACCAAGTTCTGTCAGTCCCGCAGATGGGAGGCTCCAGAAAGCAAG 709
233 IuIleValIleAspValThrLeuHisAspLeuAspAlaIleAlaValGln 249
::: ||| | | | | | | | | | | | | | | | | | | | | |
710 AGGTGTGTACACGCTGTCCCTGCGCAGATGACGATCATCACTCTGCG 759
250 ProGlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetMetLysPr 266
::: ||| | | | | | | | | | | | | | | | | | | | | |
760 ACCCAGGCG.....TTCCTGGGCGCTC.....TTCCTCAGG 788
266 cArgLysThrGluIleThrGluLysLeuArgGlnGluLysValVal 283
::: ||| | | | | | | | | | | | | | | | | | | | | |
789 TGACACAGGAGATCAAGTCAAGTCCGTGAGGAGATCAATGTCAGAG 838
283 AlAsnArgTyrIleAspGlnGlyIleAlaGluLeuValProGlyValLeu 299
::: ||| | | | | | | | | | | | | | | | | | | | | |
839 TGGCTGATGTGGCGGAGGAGGAGGAGGAGGAGATCATCTGAGTCTG 888
300 PheIleAspGluValHisMetLeuAspIleGlyCysPheSerTyrLeuAs 316
::: ||| | | | | | | | | | | | | | | | | | | | | |
889 TTCATCGACGAGGTCCACATGCTGAGCATCGAGACTTCTCTCTCA 938
316 nArgAlaLeuGluSerProIleLeuValIleLeuAlaThrAsn 333

```

seq_name:	/SIDS1/gcdata/geneseq/geneseqn/NA2000.DAT.AAA90317
seq_documentation_block:	
ID	AAA90317 standard; cDNA; 1416 BP.
XX	
XX	AAA90317;
AC	
XX	
DT	22-DEC-2000 (first entry)
XX	
DE	Yeast HEU50 DNA helicase cDNA.
XX	
KW	Yeast; HEU50; DNA helicase; TIP49 homologue; antibody; antisense therapy.
KW	diagnosis; helicase-associated disease; genetic disorder; tumour; cancer.
XX	ss.
XX	
OS	Saccharomyces cerevisiae.
XX	
Key	Location/Qualifiers
FH	1..1416
FT	CDS
FT	/*lag= a
FT	/product= "Yeast HEU50"
FT	/function= "DNA helicase"
XX	
PN	WO200047731-A1.
XX	
PD	17-AUG-2000.
XX	
PE	22-NOV-1999; 99WO-JP06519.
XX	
PR	10-FEB-1999; 99JP-0033062.
XX	
PA	(SUME) SUMITOMO ELECTRIC IND CO.
XX	
PI	Tamura T;
XX	
DR	WPI, 2000-543587/49.
XX	
RR	P-PSDB; AAB21094.
XX	

	DNA	hellicase of human origin, antibodies to iland DNA encoding it for investigation diagnosis and treatment of genetic disorders involving helicase gene expression -
Pt		
pt		
XX	Claim 8; Page 45-46; 49pp;	Japanese.
CC	This sequence represents cDNA encoding a novel yeast DNA helicase,	
CC	Heli50, which has homology with a part of human TRP49. The invention relates to human and yeast Heli50 DNA helicasases (AABZ1093), anti-	
CC	cDNAs encoding them (AAA90316, AAA90317). It also encompasses HEL50 antisense oligonucleotides and anti-HEL50 antibodies. HEL50 nucleotides, proteins and antibodies may be used for the investigation, diagnosis and CC treatment of diseases with which DNA helicase is associated, such as CC genetic disorders involving variant forms of the gene, and tumours in CC which altered levels of expression of the DNA helicase gene occur.	
XX	Sequence 1416 BP; 466 A; 238 C; 320 G; 392 T; 0 other:	
alignment_scores:		
Quality:	932.50	Length: 444
Ratio:	2.809	Gaps: 6
Percent Similarity:	74.775	Percent Identity: 44.595
alignment_block:		
US-09-589-510-4 x AAA90317 ..		
Align seg 1/1 to: AAA90317 from: 1 to: 1416		
15 llaalathrrhhsthrhislllelysgilydeungyleunspalaansglyme 31 ::: : :		
58 ATTGCTGCCACACTCCCATTATTAAGGTTTTAGTGCATGAATAAACTTGCA 107 ::: ::::		
31 talatlealeualeualaglyphvevaligylinalaalaaarglua 48 ::: ::::		
108 ACCACGTCGCCACATCCGAAGGTATGGCGCAATTGCCAACCCCGTCGTG 157 ::::		
48 laaliaglileualaialasphmetllearglnlsllysmlatalagiaryg 64 ::: ::::		
158 CTGCTGGTGTATATTTGAAATGGTTCAAATGSCACCATTAGCAGGTAGG 207 ::: ::::		
65 AlavalleuleuaallaglyproprolaalthrgilysThrAlaeualaale 81 :::		
208 GCSTTTTTGGTAGCGGGCCCCCTTCAACAAGTGAAGCACCCCTTGCCAT 257 :::		
81 uglYllaeglIngleugIyserLysValProPhecySPrometValg 98 :::		
258 GGCGITTTCCAGTCTCTGGGTMAAGATGTACCATTCTCATTTGCGG 307 :::		
98 lySergluValTySerSergluValLySlStHrGlUaLVallMeelgu 114 :::		
308 GCTCAGAANAOTTTCTTTAGAAATTGASTGAAGCTGAAGCACTAAGCA 357 :::		
115 AsnPheargarqalalieglyLeuanlgIllelySGluasnLYSLGLVAlTY 131 :::		
358 GCttTtAgGaAtTCcATGgtATCaAAATCaAGAGAGAcAAATtgAT 407 :::		
131 rglUGlIGVaItHrcGUleusErPrOgiUGLUaIgISerThThng 148 :::		
408 TGAAAGTGAAGTCTGTGSAATTCAAATTGAT.....AGATTCTATTACTG 451 :::		
148 lyclytyralalyserlieSerHisValilleleserieulystrhalv 164 :::		
452 GTGAGNCACAAACAAGCA.....AAATTGACTATTAAACTAC 489 :::		
165 lysGclYthrLysGIeulEulsleyaspSerSetlierTyASpalaleuil 181 :::		
490 GAATATGGAACAATTTATGATTAGCCAATTAATATGATTGATGCCCTAAC 539 :::		
181 elYSglULysVALiaValGILyasPVallieryILLegIUaIAAnSerg 198 :::		
540 TAAGAGAAAGCTATTGGCTGGCATGTATTCTATTGATTAAAGCTAGTG 589 :::		


```

198 1YAlaValIysArgValGlyArgCysAspSerPheAlaThrGluTyrIasp 214
    || : : : : ||||| ||||| : : ||
590 GGAGATTTCACAGCTAGCAGCA.....TCCTTCTAGATCTAGGAT 633
215 LeuGluAla.....GluGluTyrValProIleProIleGlyGlu 227
    : : : : : : : : : : : : : : : :
634 TTAGATGCCATGGCTGCTGATACCAAGATTGTTGCAATGTCGGAAGGCCA 683
227 uValHisLysLysGluIleValGlnAspValThrLeuHisAspLeuA 244
    | : : : : : : : : : : : : : : : :
684 ACTGCCAAAAGAAAGAACAGTGGTTCACACGGTGCACCTGCATGAATTG 733
244 sPaLaAlaAsnAlaGlnProGlnGlyGlnAspIleLeuSerLeuMet 260
    | : : : : : : : : : : : : : : : :
734 ATGTTATTATTCAAGACACCAAGA.....TTTTGGCATTGA... 771
261 GlyGlnMetLysProArgLysThrGluIleThrGluLysLeuArgI 277
    : : : : : : : : : : : : : : : :
772 .....TTTACTGGTGACACCGGTGAATTAGCTACAGAGCTAGAGCA 812
277 ngIuIleAsnLysValValAsnArgTyrIleAspGluGlyIleAlaGlu 294
    : : : : : : : : : : : : : : : :
813 CCAGATTAAACAAAAGTGGCAAGAGGAAGAAAGCAAGCAAGCAAA 862
294 euValProGlyValLeuPheIleAspGluValHisMetLeuAspIleGlu 310
    : : : : : : : : : : : : : : : :
863 TTGTTCTCGGTATTATTATTCGACGAAGTCCACATGTTGATATAGAA 912
311 CysPheSerTyrLeuAsnArgAlaLeuGluSerProLeuSerProIleVa 327
    | : : : : : : : : : : : : : : : :
913 TGTTTTCCCTTATTAAATAGAGCTTTGGAAAGATGATTTGCCCATTCGT 962
327 ILeLeuAlaThrAsnArgGlyIleCysAsnValArgGlyThrAspMet 344
    | : : : : : : : : : : : : : : : :
963 CATGATGCTACAAATAGAGAGCTTCCAAAACCAAGGAGCACTAATTAGA 1012
344 hnsrProHisGlyIleProValAspLeuLeuAspArgLeuValIleIle 360
    | : : : : : : : : : : : : : : : :
1013 AATCTCCACATGGGTACTCTGATCTTTGGATTAGTCAATTATTATT 1062
361 ArgThrGluThrTyrGlyProThrGluMetIleGlnIleLeuAlaIleAr 377
    | : : : : : : : : : : : : : : : :
1063 ACACTTAAAGTTATATATGACGAGAGATTAAGACAAATTTATCTATAAG 1112
377 gAlaGlnValGluGluIleAspMetAspGluGluSerLeuAlaTyrLeuG 394
    | : : : : : : : : : : : : : : : :
1113 AGCACAGAGAGGAGAGTGAAGTGCATCCGATGCCCTTGATCTATVGA 1162
394 IyGluIleGlyGlnGlnThrSerLeuArgHisAlaIleGlnLeuIleSer 410
    : : : : : : : : : : : : : : : :
1163 CCAAAACAGGTGGAAAGTATGTTGGTTACACAGTATTATTAATCTCT 1212
411 ProAlaSerValSerLysThrAsnGlyArgGluLysIleCysLysAl 427
    | : : : : : : : : : : : : : : : :
1213 GTTCTCAGCAAAATTGCATGAAGAAAGAAAACACACTGTGGAAGTGA 1262
427 aAspLeuGluGluValSerGlyLeuTyrLeuAspAlaLysSerSerAla 444
    | : : : : : : : : : : : : : : : :
1263 AGATGTCAAAAGGGCTATTATTGCTATTTTGGACAGCGCTAGATCTGTTA 1312
444 rGluLeuGlnGluGlnGlnGluLysTyrIle 454
    : : : : : : : : : : : : : : : :
1313 AGTATGTTCAAGAGACAGAGTCAACAATATATT 1344

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT.AAV84507

seq_documentation_block:

ID AAV84507 standard; DNA; 1167 BP.

XX AAV84507;

XX 01-MAR-1999 (first entry)

XX

```

DE Human secreted protein gene 97 clone HRGBR28.
XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; lissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
XX
XX MO9854963-A2.
XX
XX 10-DEC-1998.
XX
XX 04-JUN-1998; 98WO-US11422.
XX
XX 18-DEC-1997; 97US-0070923.
XX 06-JUN-1997; 97US-0048877.
XX 06-JUN-1997; 97US-0048881.
XX 06-JUN-1997; 97US-0048884.
XX 06-JUN-1997; 97US-0048893.
XX 06-JUN-1997; 97US-0048896.
XX 06-JUN-1997; 97US-0048899.
XX 06-JUN-1997; 97US-0048915.
XX 06-JUN-1997; 97US-0048949.
XX 06-JUN-1997; 97US-0048964.
XX 06-JUN-1997; 97US-0048972.
XX 06-JUN-1997; 97US-0049020.
XX 06-JUN-1997; 97US-0049375.
XX 05-SEP-1997; 97US-0057628.
XX 05-SEP-1997; 97US-0057635.
XX 05-SEP-1997; 97US-0057644.
XX 05-SEP-1997; 97US-0057647.
XX 05-SEP-1997; 97US-0057650.
XX 05-SEP-1997; 97US-0057661.
XX 05-SEP-1997; 97US-0057667.
XX 05-SEP-1997; 97US-0057761.
XX 05-SEP-1997; 97US-0057764.
XX 05-SEP-1997; 97US-0057770.
XX 05-SEP-1997; 97US-0057775.
XX 05-SEP-1997; 97US-0057778.
XX 05-SEP-1997; 97US-0048875.
XX 06-JUN-1997; 97US-0048878.
XX 06-JUN-1997; 97US-0048882.
XX 06-JUN-1997; 97US-0048885.
XX 06-JUN-1997; 97US-0048894.
XX 06-JUN-1997; 97US-0048897.
XX 06-JUN-1997; 97US-0048900.
XX 06-JUN-1997; 97US-0048916.
XX 06-JUN-1997; 97US-0048962.
XX 06-JUN-1997; 97US-0048970.
XX 06-JUN-1997; 97US-0048974.
XX 06-JUN-1997; 97US-0049373.
XX 05-SEP-1997; 97US-0057584.
XX 05-SEP-1997; 97US-0057629.
XX 05-SEP-1997; 97US-0057642.
XX 05-SEP-1997; 97US-0057645.
XX 05-SEP-1997; 97US-0057648.
XX 05-SEP-1997; 97US-0057651.
XX 05-SEP-1997; 97US-0057662.
XX 05-SEP-1997; 97US-0057668.
XX 05-SEP-1997; 97US-0057762.
XX 05-SEP-1997; 97US-0057765.
XX 05-SEP-1997; 97US-0057771.
XX 05-SEP-1997; 97US-0057776.
XX 06-JUN-1997; 97US-0048876.
XX 06-JUN-1997; 97US-0048880.
XX 06-JUN-1997; 97US-0048883.
XX 06-JUN-1997; 97US-0048892.
XX 06-JUN-1997; 97US-0048895.

```

PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA
 PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
 PI Shi Y, Soppel DR, Wei Y, Young P, Yu G, Zeng Z,
 XX
 WPI: 1999-059865/05,
 DR P-PSDB: AA088630.
 DR

Claim 4; Page 358-359; 772pp; English.

sequence 1167 BP; 245 A; 332 C; 389 G; 198 T; 3 other;

9-589-510-4 x AAV84507

2 ArgIleGluGluValGlnSerThrSerLysLysGlnArgIleAlaThrHis 18

[illegible]

```
891 CTGGCTGAGTGGCGGAGGAGGCAAGCGGAGATCATCCCTGGAGTGT 940
299 uPheIleaspGluValHisMetLeuAspIleGluCysPheSerTyrLeuA 316
|||||
941 GTTCATCGACGAGGTCCACATGCTGGACATCGAGAGGCTTCTCTCTCA 990
316 sNArgAlaLeuGluSerProLeuSerProIle 326
|||||
991 ACCGGGCCCTGGAGAGTGACATGGCGCCTGTC 1022
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 06:55:54 ; Search time 20.93 Seconds
(without alignments)
1655.969 Million cell updates/sec

Title: US-09-589-510-4

Perfect score: 2263
Sequence: 1 MRIEVOSTSKKRIATHTH.....YLDKSSARLLQEQERYIT 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR:68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query	Length	DB ID	Description
1	1723.5	76.2	456	2 JE0334	nuclear matrix pro
2	1722.5	76.1	456	2 JC5521	TATA-binding prote
3	1517.5	67.1	463	2 S52698	hypothetical prote
4	1311	57.9	458	2 T19534	hypothetical prote
5	989.5	43.7	465	2 T40697	probable tata bind
6	986.5	43.6	441	2 F75150	tbp-interacting pr
7	975	43.1	441	2 D71191	hypothetical prote
8	973	43.0	463	2 T46313	hypothetical prote
9	942.5	41.6	449	2 D69476	TBP-interacting pr
10	932.5	41.2	471	2 S61029	hypothetical prote
11	928	41.0	473	2 T46049	RuvB DNA helicase-
12	888.5	39.3	473	2 E72723	probable TATA-bind
13	863	38.1	448	2 T32710	hypothetical prote
14	299.5	13.2	112	2 T46873	hypothetical prote
15	180.5	8.0	795	2 F75154	hypothetical prote
16	179.5	7.9	798	2 B71196	probable cell divi
17	177	7.8	840	2 H71114	transitional endop
18	168	7.4	840	2 D75046	cell division cont
19	167.5	7.4	903	2 C64444	cell division cont
20	166.5	7.4	733	2 H69411	cell division cont
21	159	7.0	1251	2 B61941	hypothetical prote
22	154.5	6.8	780	1 B43859	AtPase - Sulfolobu
23	153.5	6.8	699	2 B72479	probable translati
24	150.5	6.7	627	2 S76378	cell division prot
25	148.5	6.6	726	2 D72613	probable translati
26	146.5	6.5	811	2 B69512	cell division cont
27	143	6.3	742	1 S47018	cdcl protein - Hal
28	143	6.3	742	1 C84319	cell division cycl
29	142.5	6.3	614	2 T11990	cell division prot

30	140.5	6.2	745	2 T37458	VCP-like AtPase -
31	139.5	6.2	732	2 A69086	cell division cont
32	139	6.1	464	2 G83957	probable heat-shoc
33	139	6.1	780	1 S39110	valosin-containing
34	138	6.1	809	2 T40537	AAA family AtPase
35	137.5	6.1	709	2 D81860	DNA-directed DNA p
36	136	6.0	759	2 D84301	cell division cycl
37	135.5	6.0	603	2 A84451	probable AAA-type
38	135.5	6.0	644	2 S78301	hypothetical prote
39	135.5	6.0	721	2 D82934	AtP-dependent zinc
40	135.5	6.0	835	1 S67669	transitional endop
41	134.5	5.9	655	2 E81948	probable AtP-depen
42	134.5	5.9	655	2 E81157	cell division prot
43	134	5.9	628	2 S73248	hypothetical prote
44	133.5	5.9	377	2 B70874	probable moxr prot
45	133.5	5.9	463	2 H72365	heat shock protein

ALIGNMENTS

RESULT 1
JE0334
nuclear matrix protein NMP 238 - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C:Accession: JE0334
R:Holzmann, K.; Germer, C.; Korošec, T.; Poeltl, A.; Grimm, R.; Sauerhahn, G.
Biochem. Biophys. Res. Commun. 252, 39-45, 1998
A:Title: Identification and characterization of the ubiquitously occurring nuclear ma
A:Reference number: JE0334; MUID:99032802
A:Accession: JE0334
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-456 <HOLD>
A:Cross-references: GB:A010058; NID:93892583; PIDN:CAA08986.1; PID:93892584
C:Superfamily: conserved hypothetical protein YDR190C

Query Match 76.2%; Score 1723.5; DB 2; Length 456;
Best Local Similarity 72.5%; Pred. No. 1.4e-96;
Matches 330; Conservative 73; Mismatches 51; Indels 1; Gaps 1;

```

QY 1 MRIEVOSTSKKRIATHTHIKGLDANGMAIALAGFVGQAARARAGAVDMIRK 60
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1 MKIEVKSPTKTQRIASHSHVKGIGLDESGLAKQAASGLVGQENARECGYVELISK 60
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 61 MAGRAVLLAGPATGKTALALGIAOELGSKVFCPMVGEVSEVSKTEVLMENFRRAI 120
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 61 MAGRAVLLAGPATGKTALALGIAOELGSKVFCPMVGEVSEVSKTEVLMENFRRAI 120
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 121 GLRIKKEVGEVTELSPEAEESTTGGYAKSISHVILSLKYKGTQQLKLDSSYDAL 180
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 121 GLRIKKEVGEVTELSPEAEESTTGGYAKSISHVILSLKYKGTQQLKLDSSYDAL 180
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 181 TKKVAAGVYIYEANSAGVAKRVGQDSFATPYDEAEVYVIRPGEVHKKKEIVQDTL 240
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 181 TKKVAAGVYIYEANSAGVAKRVGQDSFATPYDEAEVYVIRPGEVHKKKEIVQDTL 240
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 241 QKRVAGVYIYEANSAGVAKRVGQDSFATPYDEAEVYVIRPGEVHKKKEIVQDTL 300
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 241 QKRVAGVYIYEANSAGVAKRVGQDSFATPYDEAEVYVIRPGEVHKKKEIVQDTL 300
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 241 HDDDAANAPOGGODILSLMGOMKFRKTEYKLRQETINKVKNYRIDEGIAELVPGVLF 300
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 241 HDDDAANAPOGGODILSLMGOMKFRKTEYKLRQETINKVKNYRIDEGIAELVPGVLF 300
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 301 IDEVHMLDIECFSTLNRALESPLIVILATNKGICNVNGT-DMTSPGIPVDLLDRVI 359
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 301 IDEVHMLDIECFSTLNRALESPLIVILATNKGICNVNGT-DMTSPGIPVDLLDRVI 359
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 360 IRPFTVQPTFMIOILAIRAVERIDMDERSLAYLGEIGQTSLSRAIQLISPASVSKTN 419
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 360 IRPFTVQPTFMIOILAIRAVERIDMDERSLAYLGEIGQTSLSRAIQLISPASVSKTN 419
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 361 IRTVLYTPQEMKQIKIRATFEGINISEFALNHLGEGIKTKTLRVSVOLTPANLAKIN 420
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 361 IRTVLYTPQEMKQIKIRATFEGINISEFALNHLGEGIKTKTLRVSVOLTPANLAKIN 420
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 420 GREKICKADIEVSGVLDKSSARLLQEQERYI 454
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 420 GREKICKADIEVSGVLDKSSARLLQEQERYI 454
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Db 421 GKDSIEKEHVEISELFLYDAKSSAKIILADODQDKYM 455

RESULT 2

JC5521

TATA-binding protein-interacting protein 49 - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 02-Sep-1997 #sequence_revision 12-Sep-1997 #text_change 20-Jun-2000

C:Accession: JC5521; PC4474

R:Kanemaki, M.; Makino, Y.; Yoshida, T.; Kishimoto, T.; Koga, A.; Yamamoto, K.; Yamamoto, K.

Biochem. Biophys. Res. Commun. 235, 64-68, 1997

A:Title: Molecular cloning of a rat 49-kDa TBP-interacting protein (TIP49) that is highly

A:Reference number: JC5521; MUID:97339437

A:Accession: JC5521

A:Molecule type: mRNA

A:Residues: 1-456 <KAN1>

A:Cross-references: DDBJ:AB002406; NID:g2225876; PIDN:BA20875.1; PID:g2225877

A:Accession: PC4474

A:Molecule type: Protein

A:Residues: 1-113; 45-129-138; 446-452 <KAN2>

C:Comment: This protein plays a role in biological processes such as recombination and

C:Superfamily: conserved hypothetical protein YDR190c

C:Keywords: ATP

E:70-77/Domain: ATP-binding #status predicted <ATP>

E:302-305/Domain: ATP hydrolysis #status predicted <ATP>

Query Match 76.1%; Score 1722.5; DB 2; Length 456;

Best Local Similarity 72.3%; Pred. No. 1.6e-96;

Matches 329; Conservative 74; Mismatches 51; Indels 1; Gaps 1;

QY 1 MRIEVOSTSKORATHTHTKIGLDANGALATAAGFVGAAAEAGLAVDMIROKK 60

Db 1 MKIEVSTTKTQKQASHSHVKGIGLDESGLAQAASGLVGEANAKAGVIVELIKSKK 60

QY 61 MAGAAVLACPPATGKTALALGIAQELGSKVPCPMVGESEYSEVKTETVEMENFRRAI 120

Db 61 MAGAAVLACPPATGKTALALGIAQELGSKVPCPMVGESEYSEVKTETVEMENFRRAI 120

QY 121 GLRKEKEVEYEGEVELSPKEASTTGGAASISHVITSLKTVGKTKOLKIDSSITDAL 180

Db 121 GLRKEKEVEYEGEVELSPKEASTTGGAASISHVITSLKTVGKTKOLKIDSSITDAL 180

QY 121 GLRKEKEVEYEGEVELSPKEASTTGGAASISHVITSLKTVGKTKOLKIDSSITDAL 180

Db 121 GLRKEKEVEYEGEVELSPKEASTTGGAASISHVITSLKTVGKTKOLKIDSSITDAL 180

QY 181 IKKRVAVGVYIYIANSAGVAVKRVGCDSPATEDELEAEVYIPKGEVKKKEIYODVTL 240

Db 181 IKKRVAVGVYIYIANSAGVAVKRVGCDSPATEDELEAEVYIPKGEVKKKEIYODVTL 240

QY 241 HDLDAAANQPOGGODILSLMGOMKPRKTEITKLRQETINNVNRYIDEGIAELVPGVLF 300

Db 241 HDLDAAANQPOGGODILSLMGOMKPRKTEITKLRQETINNVNRYIDEGIAELVPGVLF 300

QY 301 IDEVHMLDIECFSYLNRALSPSPVILATNRGICNVRGT-DMTSPHGIPLVDLRLVI 359

Db 301 IDEVHMLDIECFSYLNRALSPSPVILATNRGICNVRGT-DMTSPHGIPLVDLRLVI 359

QY 360 IRTETYGPEMIIQILAIRAQVEIDMDESLAYLIGEOQTSLRHAQLISPAVSKTN 419

Db 360 IRTETYGPEMIIQILAIRAQVEIDMDESLAYLIGEOQTSLRHAQLISPAVSKTN 419

QY 420 GREKICKADLEFVSGTLYLDKASSARLLDOEQERYI 454

Db 420 GREKICKADLEFVSGTLYLDKASSARLLDOEQERYI 454

QY 421 GKDSIEKEHVEISELFLYDAKSSAKIILADODQDKYM 455

RESULT 3

S52698

hypothetical protein YDR190c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YD9346.02c

C:Species: Saccharomyces cerevisiae

C>Date: 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999

C:Accession: S52698

R:Oliver, K.; Harris, D.

submitted to the EMBL Data Library, March 1995

A:Reference number: S52698

A:Accession: S52698

A:Molecule type: DNA

A:Residues: 1-463 <OLL>

A:Cross-references: EMBL:Z48784; NID:9755782; PIDN:CAA8704.1; PID:g755784; GSPDB:GNO

C:Genetics: A:Experimental source: strain AB972

A:Gene: MIPS:YDR190c

A:Map position: 4R

C:Superfamily: conserved hypothetical protein YDR190c

Query Match 67.1%; Score 1517.5; DB 2; Length 463;

Best Local Similarity 65.7%; Pred. No. 3.6e-84;

Matches 286; Conservative 81; Mismatches 67; Indels 1; Gaps 1;

QY 14 RIATHTHTKIGLDANGALATAAGFVGAAAEAGLAVDMIROKKMGAAVLALGPPA 73

Db 23 RIATHTHTKIGLDANGALATAAGFVGAAAEAGLAVDMIROKKMGAAVLALGPPA 73

QY 74 TGTALALGIAQELGSKVPCPMVGESEYSEVKTETVEMENFRRAIIGRIKKEVYEG 133

Db 83 TGTALALGIAQELGSKVPCPMVGESEYSEVKTETVEMENFRRAIIGRIKKEVYEG 133

QY 134 EYTELSPKEASTTGGAASISHVITSLKTVGKTKOLKIDSSITDALIKKAVGDDVYI 193

Db 143 EYTELSPKEASTTGGAASISHVITSLKTVGKTKOLKIDSSITDALIKKAVGDDVYI 193

QY 194 EANSAGVAVKRVGCDSPATEDELEAEVYIPKGEVKKKEIYODVTLHDLDAAANQPOGG 253

Db 203 EANSAGVAVKRVGCDSPATEDELEAEVYIPKGEVKKKEIYODVTLHDLDAAANQPOGG 253

QY 254 ODILSLMGOMKPRKTEITKLRQETINNVNRYIDEGIAELVPGVLFIDEVHMLDIECF 313

Db 263 ODILSLMGOMKPRKTEITKLRQETINNVNRYIDEGIAELVPGVLFIDEVHMLDIECF 313

QY 314 YLNRALSPSPVILATNRGICNVRGT-DMTSPHGIPLVDLRLVIIRTEYGTETIQ 372

Db 323 YLNRALSPSPVILATNRGICNVRGT-DMTSPHGIPLVDLRLVIIRTEYGTETIQ 372

QY 373 ILAIRAQVEIDMDESLAYLIGEOQTSLRHAQLISPAVSKTNREKRIEVEV 432

Db 383 ILAIRAQVEIDMDESLAYLIGEOQTSLRHAQLISPAVSKTNREKRIEVEV 432

QY 433 SGLYLDKASSARLLDOEQERYI 454

Db 443 SGLYLDKASSARLLDOEQERYI 454

RESULT 4

T19534

hypothetical protein C27H6.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C:Accession: T19534

R:Gardner, A.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19138

A:Accession: T19534

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-458 <WIL>

A:Cross-references: EMBL:Z81042; PIDN:CAB02793.1; GSPDB:GN00023; CESP:C27H6.2

C:Genetics: A:Experimental source: clone C27H6

A:Gene: CESP:C27H6.2

A:Map position: 5

A:Introns: 22/3; 54/3; 292/1; 413/1; 426/3

C:Superfamily: conserved hypothetical protein YDR190c

Query Match 57.9%; Score 1311; DB 2; Length 458;

Best Local Similarity 57.8%; Pred. No. 9.7e-72;

[illegible][illegible]

OY 419 NREKICKADLEEVSGLYIDAKSSARLLQOE 448
 Db 407 -GGKREHREHVEKAKEFYADVKRSIAFVEK 435

RESULT 7

hypothetical protein PH1804 - Pyrococcus horikoshii
 C:Species: Pyrococcus horikoshii
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
 C:Accession: D71191
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hatakeyama, Y.; Hino, Y.; Yamamoto, S.; Seki, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
 A:Reference number: A71000; MUID:98344137
 A:Accession: D71191
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-441 <RAW>
 A:Cross-references: GB:AP000007; MID:93236134; PIDN:BA30923.1; PID:93258240
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Genetics:
 A:Gene: PH1804
 C:Superfamily: conserved hypothetical protein YDR190C

Query Match

Best Local Similarity 43.1%; Score 975; DB 2; Length 441;
 Matches 207; Conservative 92; Mismatches 131; Indels 18; Gaps 6;

OY 3 IEEVOSTSKORATHTHTKGLDANGMAIALAGFVGOAAREAGLAVDMIROKMA 62
 Db 4 IEEPLAL-KERYGAHSHIRKGLIDENGAKRFIGDMGVYARABAGIAVKIKOGSLA 62
 OY 63 GRAVLLAGPPATGKTALALGIAOELGSKVPCPMVGEVSEVSKTEVLMENFRRAIGL 122
 Db 63 GKILLVGPSTKTAIANGIAIRELGEDEVPFOISGEYSAEVKKTEFLQALFRAIGV 122
 OY 123 RIKENKEVEGEVTELSPEAEESTTGGYAKSISHYIISLKTIVKGTOKLDSIYDALI 182
 Db 123 RISEEKVVEGEVKEVKEIKRTHPNPIELPESIRITLTKDKDKTIRAGRELAVYOLD 182
 OY 183 EKAVADVIYEANGAVKRVGRCDSPATEVDLEAEVVPJPKGEVHKKEIVODVTLAD 242
 Db 183 MEVEGDVYQIDAEETGRVSKIG--TTKEEGLEFKKKVELPTGPVLKIKETFTVTLHD 239
 OY 243 IDAANAPOGGODILSLM--GQMKPRKTEITKLRQEVNRYIDEGLAELVPGVLF 300
 Db 240 IDVNAARAG--IFSLIFGGM-----ELNDEIRENVDTGVAKQMLEBKATLVPGVLF 290
 OY 301 IDEVHMLIECESYINRALESPLSYIATLNRCINAVRGDTMTSPHGIPVLDLRLVIT 360
 Db 291 IDECHMLIEAFSFLARMENELAPILILATNRGKTRIGTDIEAPHGIPVMDLRLIIT 350
 OY 361 RTETVPTMETQIOLAIRAOVEIDMEESLAYIGEIGQOSLSRAHOLISPAVSVKTN 420
 Db 361 NTEPKKOEIIRIKIRAKKEVLESEALEYIADLGEKTSLSRYAVOLLAPASIIA---G 407
 OY 421 REKICKADLEEVSGLYIDAKSSARLLQOE 448
 Db 408 GKREKEHREHVEKAREYFADIKRSISFEVK 435

RESULT 8

T46313
 hypothetical protein DKFZp434K1011.1 - human
 C:Species: Homo sapiens (man)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Feb-2000
 C:Accession: T46313
 R:Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Well, B.; Wiemann, S.

submitted to the Protein Sequence Database, January 2000
 A:Reference number: 223036
 A:Accession: T46313
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-463 <AAU>

A:Cross-references: EMBL:AL136743
 A:Experimental source: adult testis; clone DKFZp434K1011
 C:Genetics:
 A:Note: DKFZp434K1011.1
 C:Superfamily: conserved hypothetical protein YDR190C

Query Match

Best Local Similarity 43.0%; Score 973; DB 2; Length 463;
 Matches 200; Conservative 97; Mismatches 144; Indels 14; Gaps 4;

OY 2 RIEVOSTSKORATHTHTKGLDANGMAIALAGFVGOAAREAGLAVDMIROKMA 61
 Db 9 KVEIRVOTRIERIGAHSHIRKGLIDALEPRQASOGMVOLAARAAVLEMIREGKI 68
 OY 62 GRAVLLAGPPATGKTALALGIAOELGSKVPCPMVGEVSEVSKTEVLMENFRRAIG 121
 Db 69 AGRAVLLAGPGTGKTAIANGMAALGPDPTPTAIGSEIFLSKTEALTQAFRRSIG 128
 OY 122 LRIKENKEVEGEVTELSPEAEESTTGGYAKSISHYIISLKTIVKGTOKLDSIYDALI 181
 Db 129 VRIKEETELIEGEVVEIYQIDPRATGTSKVGK-----LTKETMETTYIDGTMIESLT 183
 OY 182 KERYAVADVIYEANGAVKRVGRCDSPATEVDLEAE--YPIPKGEVHKKEIVODVT 239
 Db 184 KDVVQAGDVITIDKATGKISKLRSTFRADYDAMGSOQKVFQCPDGLQRRKEVHVTS 243
 OY 240 LHDLDAAANPOGGODILSLMOMKPRKTEITKLRQEVNRYIDEGLAELVPGVLF 299
 Db 244 LHEIDIVINSRTG--FLAL----FSGDGEIKSEVRQIANAKVAEMREKATILPGVL 296
 OY 300 FIDEVHMLIECESYINRALESPLSYIATLNRCINAVRGDTMTSPHGIPVLDLRLVIT 359
 Db 297 FIDEVHMLIECESYINRALESPLSYIATLNRCINAVRGDTMTSPHGIPVLDLRLVIT 356
 OY 360 IRTETVPTMETQIOLAIRAOVEIDMEESLAYIGEIGQOSLSRAHOLISPAVSVKTN 419
 Db 357 VSTTPYSEKDTQOLIRICEEDVMSDATVTLRGLETSLRYATOLITAAASLVCRR 416
 OY 420 GREKICKADLEEVSGLYIDAKSSARLLQOEQERYT 454
 Db 417 KGEVQVDIDIKRVYSILFIDESRSTQYMKRYODAFI 451

RESULT 9

DBP-interacting protein TTP49 homolog - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
 C:Accession: D69476
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
 G. Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirness, E
 Nature 390, 364-370, 1997
 A:Authors: Uitterlinden, T.; Cotton, M.D.; Springs, T.; Arltach, P.; Kaine, B.P.; Sykes,
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
 A:Reference number: A69250; MUID:98049343
 A:Accession: D69476
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-449 <MLE>
 A:Cross-references: GB:AF000977; GB:AF000782; MID:92689300; PIDN:AB89434.1; PID:9264
 C:Superfamily: conserved hypothetical protein YDR190C

Query Match

41.6%; Score 942.5; DB 2; Length 449;

RESULT 12
 E72723
 Probable TATA-binding protein-interracting protein 49 APE0326 - Aeropyrum pernix (strain C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: E72723
 R:Kavaraavasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
 A:Reference number: A72450; MUID:99310339
 A:Accession: E72723
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-473 <KAW>
 A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BA479281.1; PID:95103965
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE0326
 C:Superfamily: conserved hypothetical protein YDR190C

RESULT 13
T32710
hypothetical protein T22D1.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T32710
R:Geisels, C.; Bradshaw, H.; Hawkins, M.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid T22D1.
A:Reference number: Z21211
A:Accession: T32710
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-448 <GEI>
A:Cross-references: EMBL:AF039052; PDB:AA94278.1; GSPDB:GM00022; CESP:T22D1.10
A:Experimental source: strain Bristol N2, clone T22D1
C:Genetics:
A:Gene: CESP:T22D1.10
A:Map position: 4
A:Introns: 14/3; 60/3; 103/3; 196/1; 326/2; 377/2
C:Superfamily: conserved hypothetical protein YDR90C

Query Match	38.1%	Score 863;	DB 2;	Length 448;
Best Local Similarity	42.2%;	Pred. No. 9e-45;		
Matches 191;	Conservative 83;	Mismatches 161;	Indels 18;	Gaps
QY	6	VOSTSKORATHTHIKGLDANGMAIALAAAFVQAAAREAGLAVDMIROKKWA	65	
Db	9	VKDIIVKERTSVSHSTIGTGLDRLREAEVYSGSGMVAAARAAGLIYVMIKGTAK	68	
QY	66	VILASPANGKTLALAGIAGDELGSKPPCPGMYGEVYSEKVTGYLMMFRRAIG	125	
Db	69	LVTGTEGMAKTAIALAISKELGEDNPPEFVISAASEIYMEINKTEALTQAFRA	128	
QY	126	ENKEYEGEETELSPREASTTGVYAKSISHYIISLTKYGTOKLLOSSIYDALIK	185	
Db	129	EETEVELEGEVISL---EYDRSANGMGPKVGK--LMMRTDMETIYDLGSKMAD	183	
QY	186	AVGDVYIIEANGSAVVRVRCDSFATEYDLEA---EEVYIPKGEVHKKEIVD	241	
Db	184	MEDVYIQVKKASGRVTRIGR--SEKRSHDYDAMGPKVYVQCDPDEIQKRRE	241	
QY	242	DLDANAPOGQODILSLMGOMMKPRKTEITEKLEQETINKVYNRTIDGIELV	301	
Db	242	DIDIVISFRQGYALFS-----GDTGEIKAEVRDQINKKYLEKREGGKAK	294	
QY	302	DEVHMLDIECESYLNRALLESPLSVIILATNRGICNVRCGDMTSPHGIV	361	
Db	295	DEAHMLDIECESFLRALEGESPLIMATNRLIEKVGTVGESAHGIPSPDL	354	
QY	362	TETVPTEMIQIATRAVEELDMDEESLAVGEGOOTSRHMLTOLISPASVSK	421	
Db	355	ALPYTKEDTAKLISRQDEEGVKIQPTALDLVLKQETISLRTCITHLTMA	414	
QY	422	EKICADILEEVSGLYLDAKSSARLLQEOERYI	454	
Db	415	EIVYTDHGSAYRLEFDTRKSEKILTESAGFL	447	

Wed Nov 14 08:34:17 2001

us-09-589-510-4.std.rpr

Db	214	KGVLIVGPPTGKTLIAKANVANDAGN	--	FVIYNPEIMSKYGTGE	----	ENLAK	-IFEE	267		
Oy	124	IKENKE	--	VEGEVTELSPEEAEESTGTGA	KSISHVILSLTKVGTOL	-----	K	177		
Db	268	AEEADPSIIF	IIDIDIAIAKRD	EAETGEVERRLVAOLLITMDGLKGRGVV	IGATNPNM	327				
Oy	172	LDSSI	----	YDALIKEKAVG	-----	DYIY	-----	193		
Db	328	LDPALRRGRFPD	----	RELIVGPPREGREKELQIHTRN	MLADVDLDIADVTHTGFCV	383				
Oy	194	----	----	EANGAYKRGRCDSFATEYDL	EAEEVPIIPKGEVKKREIVODVTL	HLHDA	245			
Db	384	ADLAALCKEAMARALRRV	----	LP	SIDLEAEE	--	IPKEVDNLK	-----	VTMDFK	429
Oy	246	A	--	MAOPGGODIL	-----	SLMGOMKRRKTEIEKLR	OEIN	-----		280
Db	430	ALKDVEPSAMREVL	VEPVNK	WEDIGL	----	EEVQOE	LEAEWPLKAEVEK	IGVR	484	
Oy	281	----	----	RVNRYIDE	----	GAEL	-----		294	
Db	485	PRGVLLGPPGTGKTLIAKANV	ANESGANFISV	KGPEIFSKWNG	GESKALREIF	PARAOS	544			
Oy	295	VGVLFEID	EVHMLDIECP	SYLNR	-----	LESPLSPVILATN	RGICNR	339		
Db	545	ACCIIFPEIDAIAPKRGR	OLSSAVTDKVNQ	LTTELDGNEEKRDV	VLAATNR	-----		598		
Oy	340	GTDMTSPG	ITVDLLRLVI	-----	IRTEYTGPEMIQ	--	LAIR	-----	377	
Db	599	PDIIIDPALLRGR	LDIVILV	IPVDEKARLDIFKIH	RRSMNLAEADVNLEELAKK	IEGTG	657			
Oy	378	AOVEIDIMDESL	YALGEIO	----	QUSPHAL	--	QUSPA	-----	SVSKTNGR	421
Db	658	ADIEALCREAMLA	VRISIKRPM	WIDIEKLEL	ELINLYQ	OSIGTFRAAAV	ELNSV	IKATKER	717	
Oy	422	EKICADLEYS	----	GLYIDAKSSAR	LLOEOQERYI	454				
Db	718	ESABAGEFSELKNA	IGITIVLS	PSAPEKIEA	VEVKEI	753				

RESULT	2
SAV_SULIAC	
ID	SAV_SULIAC
AC	007590;
DT	01-FEB-1995 (Rel. 31, Created)
DT	01-FEB-1995 (Rel. 31, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
GN	SAV PROTEIN.
DE	
OS	Sulfolobus acidocaldarius.
OC	Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX	NCBI_TaxID=2285;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-ATCC 33909 / NCIB 11770 / DSM 639;
RX	MEDLINE=94118286; PubMed=8289263;
RA	Confalonieri F., Marsault J., Duguet M.;
RT	"SAV, an archaeobacterial gene with extensive homology to a family of
RL	highly conserved eukaryotic ATPases.";
J.	Mol. Biol. 235:396-401(1994).
-!	FUNCTION: NOT YET KNOWN, SHOWS ATPASE ACTIVITY.
-!	SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES. p97/CDC48
CC	SUBFAMILY.
CC	
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: L17042; AAA72002.1; -
DR	InterPro: IPR001939; -.

DR Pfam: PF00004; AAA, 2.
DR PROSITE: PS00674; AAA, 2.
KW ATP-binding; Repeat.
FT NP_BIND 253 260 ATP (POTENTIAL).
FT NP_BIND 528 535 ATP (POTENTIAL).
SQ SEQUENCE 780 AA; 87480 MW; 669B2846BD05E45E CRC64;

Query Match	6.88;	Score 154.5;	DB 1;	Length 780;
Best Local Similarity	22.88;	Pred. No. 0.039;		
Matches 97; Conservative	69;	Mismatches 154;	Indels 105;	Gaps 22

```

0Y      64 RAVLWAGPARGKALMALGIAOELSKVPCFWMGSEVYSEVKTVEY-LMENFRRAIGL 12
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     247 KGIILXGPPGTGKTLRALRNEIGAY--FITVNGPELIMSIFYESESORIREIFKEA--- 30
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     123 RIKENKE--VEYEGVTELSPEEAESTYGGYAKSISHVILSKT-VKGTOKLIKDSIYDA 17
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     302 --EENAPSIIFIDIBIDALARK-EDVTGYEKRYVAOLLTMOJINGRGVY----- 34
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     180 LKEKENVANGDYIYEANSQAVKRVKRGCSFATEYDELEAEVYDIPKG---EVH-KKKE 23
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     350 ----IYIGATNRPDADIPALRRPGR-----FDREIIRPPDTGGRDILQVHTRRMP 39
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     234 IVODVTLHDDANAPPOG--GODLISL-----NGOMAKRRK-----TEITEK 27
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     398 ITDDV---DLDIKAEWTYGYTADIALAKAFAIYALRRRVDEKKNLMDPTIPAEIIKE 45
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     275 LRQIEINKVNRKYIDESIAELVGVL-----FIDEVHMLDI---ECFSYLNALLESPISP 32
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     455 LKVSMDNDELN-----ALKSIQPSLLREYYVEVRVNMNDIGGLDNVKKOOLRRAVEMPLR- 50
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     326 IVILATNRGICNVRGDMSPHGIPVDLDLRYL-----IKRETYGPTEMI 37
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     509 FPELFTSSGVTTPKGGILLGPPGTGKTMKLKAAVATESGANFIYAVGPETLSKMWGESEK- 56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     372 QILAIRAQVEEIDIMDESILAYIGEIGOTSIRHAIULISPAVSKTNG-REKICAKDLE 43
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     568 ---AIRIEFRKARQAAPVYIFDEIDS-----IAPINGLSTDSGYTERIVNOLLA 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y     431 EWSGL 435
      : : : :
Db     615 EMDGI 619

```

	RESULT	3
	FTHL_SYN3	
ID	FTHL_SYN3	STANDARD; PRT; 627 AA.
AC	Q55700;	
DT	01-NOV-1997 (Rel. 35, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	01-NOV-1997 (Rel. 35, Last annotation update)	
DE	CELL DIVISION PROTEIN FTSH HOMOLOG 1 (EC 3.4.24.-), SLR0228.	
OS	Synechocystis sp. (strain PCC 6803).	
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.	
OX	NCBI_TaxID=1148;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=96127529; Pubmed=8590279;	
RA	Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N., Suda Y., Tabata S.;	
RT	"Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome."	
RL	DNA Res. 2:153-166(1995).	
CC	-1- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE (BY SIMILARITY).	
CC	-1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).	
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.	
CC	-1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41 (ZINC METALLOPROTEASE).	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: D64000; BAI0230.1; -
DR MEROPS: M1.005; -
DR Interpro: IPR000130; -
DR Interpro: IPR000642; -
DR Interpro: IPR001939; -
DR Pfam: PF00004; AAA; 1.
DR Pfam: PF01434; Peptidase_M1; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00674; AAA; 1.
KW Cell division: ATP-binding; Transmembrane; Hydrolyase; Metalloprotease;
KW Zinc; Multigene family.
KW DOMAIN 1 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 POTENTIAL.
FT DOMAIN 29 117 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 118 138 POTENTIAL.
FT DOMAIN 139 627 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 212 219 ATP (POTENTIAL).
FT METAL 433 433 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 434 434 BY SIMILARITY.
FT METAL 437 437 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 627 AA; 68496 MW; 4B2C160213CD0517 CRC64;

Query Match 6.7%; Score 150.5; DB 1; Length 627;
Best Local Similarity 21.0%; Pred. No. 0.051;
Matches 105; Conservative 75; Mismatches 159; Indels 161; Gaps 26;

QY 25 GLDNGMAIALAAGFVQAARPAAGLAIDMIRKMAAGRAVLAAPATGTALALGIA 84
DB 176 GIDAKKELEQVYFLKOPERTAVGAKIP- - - - - KGVLLVPPPTGTTLAKAIA 226
QY 85 QELSGKVPFCPMVSGE- - - - - VYSEVKTTEVLMENFRRAIGLRIKENKE- - - - - VYEGEV 135
DB 227 GEAG- - - - - VPFSSISGEVEFVGASRVR- - - - - DLFKKA- - - - - KENAPCLIFIDRI 273
QY 136 TELSPREASTTGG- - - - - YAKSISHVILSLKTVKQTKQLKDSITYALLIKEVAVGVLY 192
DB 274 DAVGQRGAGIGGNDREDEQTLNOLTEMDFEGN- - - - - TGIIT 314
QY 193 IEAN- - - - - SGAVKRVGRCDSPATEYDLAEVVPKPG- - - - - EVH- - - - - KKEIVQDVTLH 241
DB 315 AATNPDPVLDALMPGR- - - - - FDRQVWADAPDY- - - - - SGKELLEVARMKKILAPEVSI 366
QY 242 LIDANAOPOGODILSLMG- - - - - MMKPKTEITELKROELINKVNRITD- - - - - EG- - - - - 290
DB 367 DSIAKRTGFGSADLANLNEALITLARRKSAIT- - - - - LEIIDAADVORVAVAGMEGPLVD 423
QY 291 - - - - - IAEVPGVLFIDEVHMLDIECFSYLNRALESPLSPVILATNRGICNVR 339
DB 424 SKSKRLIAVHEVGHAVITLLKD- - - - - HDVQKTLIP- - - - - RG- - - - - QAQ 462
QY 340 GTDMSPHGIPVDLDRVITRTETVPTMIQILAI- - - - - RAQVEIDMDE- - - - - 388
DB 463 GLTWTFPN- - - - - EEOGLTTRKAOLMARITAGAMGRAAEVEFGDDEVTG 507
QY 389 - - - - - SLAVIGEIQOQTSLRHAIDILSPASVSKTN- - - - - GRKIKADLEEVSGLYLDA- - - - - 439
DB 508 AGGDILQOQVEMAROWITRGMSNLGPILESSEGVFLVGGILMNRSEI SEVATRIDAOV 567
QY 440 - - - - - KSSARLLQEOOE 451
DB 568 ROLAEOGHOMAKIYDQRE 587

RESULT 4
CDCH_HALNI STANDARD: PRT: 742 AA.
AC Q9HBF0;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CDCH PROTEIN
GN CDCH OR CDC48C OR VNG1667G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;

RA [1]
RP MEDLINE-20504483; PubMed-11016950;
RA Ng W.V., Kennedy S.P., Maharias G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weller R., Goo Y.A.,
RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RA "Genome sequence of Halobacterium species NRC-1."
Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RL -1- FUNCTION: MAY BE PART OF A TRANSDUCTION PATHWAY CONNECTING LIGHT
-1- TO CELL DIVISION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES. P97/CDC48
SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE005075; AAG19919.1; -
DR PROSITE: PS00674; AAA; 2.
KW ATP-binding; Repeat.
FT NP_BIND 230 237 ATP (POTENTIAL).
FT NP_BIND 503 510 ATP (POTENTIAL).
SQ SEQUENCE 742 AA; 81657 MW; 4E248D2F87ABE126 CRC64;

Query Match 6.3%; Score 143; DB 1; Length 742;
Best Local Similarity 19.6%; Pred. No. 0.18;
Matches 90; Conservative 74; Mismatches 151; Indels 144; Gaps 18;

QY 15 IATHTHKGIGLIDANGMAIAL- - - - - AAGFV- - - - - GQAARPAAGLAIDMIRKMA- - - - - 62
DB 151 IAVETEPGVCLVEDTDELREPISGFERTGGITIEDIGLENIQVRKVELPK 210
QY 63 - - - - - GRAVILAGPATGKTALALGIAQELSGKVPFCPMVSGEVEYSEVKTTE 110
DB 211 HPQIFKQIGLEPPGVLLHGPRTGKTLAKAVANE- - - - - TSASFPSIAGPEILISKYGESE 268
QY 111 -VLMENFRRAIGLRIKENKE- - - - - VYEGEVTLSPEASTTGGYAKSISHVILSLKTVKGT 167
DB 269 QQLRELFEDA- - - - - KDSPSILIFIDEIDSTAPKREDTGEVERRVAAQLITMDGIEGR 323
QY 168 KQIKLDSITYDLLEKAVAGGVITYEANSKAVKRVGRCDSPATEYDLAEVVPKPG- 226
DB 324 GOV- - - - - VTAATNVDAVDALRRPGR- - - - - PRELEITGVPEIDIGR 362
QY 227 - - - - - EVHKK-KEIVQDVTLHDIDANAOPOGODILSLMGOMKPKTEITELKROELINK 281
DB 363 EELIKHTRGMPILSDVDNLT- - - - - ADDTGFGVADIESLS- - - - - KEAAMR 406
QY 282 VNRYTDEGIAELVPGVLFIDEVHMLDIECFSYLNRALESPLSPVILATNRGICNVRGT 341

```

Db 407 ALRRYLPE-----IDLE-----419
QY 342 DMTSPHIGIVDLDRLVIIRTEYGTPTMIOQLAIR---AQVEIDMDESLAYLGEIQ 398
Db 420 -----EDIPSLIDRMIVKREDKALSEVEPSARREVLVEPKITWD-----DVG 466
QY 399 QTSLRHAIOQLISPAVSXKNGREKICKADLEEVSGLY 437
Db 467 LTEAKNNVK-----ESVWPLNQPEKFTMGVEPPAGVLL 501

RESULT 5
CDCH_HALSA STANDARD: PRT: 742 AA.
AC P46464;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CDCH PROTEIN.
GN CDCH.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
NCBI_TaxID=2242;
RN NCBI
RP SEQUENCE FROM N.A.
RC STRAIN=RM1;
RL Bibikov S.I., Oesterhelt D.;
Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: MAY BE PART OF A TRANSDUCTION PATHWAY CONNECTING LIGHT
TO CELL DIVISION.
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES. P97/CDC48
SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X79560; CAA56097.1; -
DR InterPro; IPR001939; -
DR Pfam; PF00004; AAA; 2.
DR PROSITE; PS00674; AAA; 2.
KW ATP-binding; Repeat.
FT NP_BIND 230 237 ATP (POTENTIAL).
FT NP_BIND 503 510 ATP (POTENTIAL).
SQ SEQUENCE 742 AA; 81644 MW; A1107D3F48B51F18 CRC64;

Query Match 6.3%; Score 143; DB 1; Length 742;
Best Local Similarity 19.6%; Pred. No. 0.18;
Matches 90; Conservative 74; Mismatches 151; Indels 144; Gaps 18;

QY 15 IATHTHKGIGLDANGMAIAL---AAGFV---GQAAAREAGLAVDIRKKMA-----62
Db 151 IAVETPEEGVCLVETDTEDELNEEPIISGERGGITTEDIGGLENEQREWEVLPK 210
QY 63 -----GRAVLAGPATKTLALGIAOELGSKVFCPMVGSEVSEVKT 110
Db 211 HQIIFOKIGIEPPGVLLHGPETGKTLARAVANE--TSASFSGIAPETISYGESE 268
QY 111 -VIMENFRRAIGLRKENKE--VYEGEVTLESPBEASTGTGAKSISHYISLKYKGT 167
Db 269 OQLREIFEDA-----KDSPSIIFDELDSIAFKREDTGEVRRVVAOGLLTMDGLEG 323
QY 168 KQLKLDSSITDALKEKAVAGDVYIEANSQAVKRGCSFATYEDLEAEVYPIPKG- 226
Db 324 GGV-----IVIAATNRVAVDPALRRPGR-----FDRELEIGVPPDEIGR 362
QY 227 -----EVHKK-KEIVQDVTLHLDDAANAQPOGGDILSLMCOMMKPKRTEITERKLEIRNK 281

```

```

Db 363 EELIKHTRMPSLSDVYNSTL-ADDTGCFVADIESIS-----KEAMR 406
QY 282 VVNRKIDEGIAELVPGVLFIDEVHMLDECFSYLNRALPSLPVILATNRGICNVRGT 341
Db 407 ALRRYLPE-----IDLE-----419
QY 342 DMTSPHIGIVDLDRLVIIRTEYGTPTMIOQLAIR---AQVEIDMDESLAYLGEIQ 398
Db 420 -----EDIPSLIDRMIVKREDKALSEVEPSARREVLVEPKITWD-----DVG 466
QY 399 QTSLRHAIOQLISPAVSXKNGREKICKADLEEVSGLY 437
Db 467 LTEAKNNVK-----ESVWPLNQPEKFTMGVEPPAGVLL 501

RESULT 6
FTTH_HAETN STANDARD: PRT: 635 AA.
AC P71377;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELL DIVISION PROTEIN FTSH HOMOLOG 1 (EC 3.4.24.-).
GN FTSH-A OR H11335.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
NCBI_TaxID=727;
RN NCBI
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Kelschman R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervase A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA Mcmeney K., Sutton G., Fitzhugh W., Fields C.A., Goocyne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Wetman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uetback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geophagen N.S.M.,
RA Gnehm C.U., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kervase A.R., Fleischmann R.D.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE
(BY SIMILARITY).
CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41 (ZINC
METALLOPROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U32812; AAC22979.1; -
DR TIGR; H11335; -
DR InterPro; IPR000642; -
DR InterPro; IPR001939; -
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF01434; Peptidase_M41; 1.
DR PROSITE; PS00674; AAA; 1.
KW Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;

```


Query Match
Best Local Similarity 23.9%; Score 142; DB 1; Length 635;
Matches 85; Conservative 51; Mismatches 116; Indels 104; Gaps 21;

37 AGFVGAARAGLAVDMIRQ---KMG---RAVLAPGATKTLALGIAOELGS 89
151 ADVAGCEAKKEVEIVDFLDPNPKFONLGKIPKGLMVGPGTGKTLARAIAGE--A 208

90 KVPCCPMWSE-----VYSEVKTEVLMENFRRAIGLRKENKE--VYEGEVELSP 140
209 KVPFETISGFVEMFVGASRYR-----DMFEQA-----KKNAPCLIFIDEIDAVGR 257

141 BEAETTGCT---AKSISHYISLKTGKTKOLKLDSSYDALIKEKVAAGVYIEANS 197
258 GAGAGIGGHEROTLNQMIVEMDGFSGN-----DGVYIAATNRP---DVL-----D 303

198 GAVKRGCDSPATEYDLAEYVPIPG-----EVHKRR-ETVODVTLHDLDAANAPOG 251
304 PALTRPR-----FDRQVVGGLPDYGVKREQLIKVHKRYVAQDVADMTL-ARGTPGT 355

252 GGDILSLMGQ---MMKPKTEIT---EKLROEIN-----KVNRYIDE 289
356 SGADILNVEALPAARANKRTVMTLEFEKAKDKINKGPRRRMTMTDKOKETAHHEA 415

290 G---IAELIPGVLFIDEVHMDI-----ECFSYLNRALSPLSP 326
416 GHAIVGYLVE---HDPVHKVTIIPRGALGVTFPLPEGDISISOKLESKSTL 468

RESULT 7
VAT_THEAC STANDARD; PRT; 745 AA.

AC 005209;
DT 01-OCT-2000 (rel. 40, Created)
DT 01-OCT-2000 (rel. 40, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE VCP-Like ATPase.
GN VAT OR TA0840.
OS Thermoplasma acidophilum.
OC Archaea: Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma
OX NCBI_taxid=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=97227958; PubMed=9119075;
RA Baumeister W., Tamura T., Lupas A.N., Peters J., Cejka Z., Ashraf W.,
Baumeister W.,
"Cloning, sequencing and expression of VAT, a CDC48/p97 ATPase
homologue from the archaeon Thermoplasma acidophilum.",
FEBS Lett. 404:263-268(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Korelke K.K., Volker C.,
Mewes H.-W., Fritsman D., Stocker S., Lupas A.N., Baumeister W.,
"The genome sequence of the thermophilic scavenger Thermoplasma
acidophilum",
Nature 407:508-513(2000).

3D-STRUCTURE BY ELECTRON TOMOGRAPHY.
RP STRAIN=DSM 1728;
RX MEDLINE=99283870; PubMed=10356978;
RA Rocco B., Walz J., Hegert R., Peters J., Baumeister W.,
"Structure of VAT, a CDC48/p97 ATPase homologue from the archaeon
Thermoplasma acidophilum as studied by electron tomography.",
FEBS Lett. 451:27-32(1999).
RL FUNCTION: HAS AN OPTIMAL MG2(+)-ATPASE ACTIVITY AT 70 DEGREES
CELSIUS.
CC SUBUNIT: HOMODIMER. FORMS A RING-SHAPED PARTICLE.
CC SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES. P97/CDC48
SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.stb.ch)
CC
CC EMBL: U78072; AAC45089.1;
CC EMBL: AL445065; CAC11969.1;
CC InterPro: IPR001939;
CC Pfam: PF00004; AAA; 2.
CC PROSITE: PS00674; AAA; 2.
CC ATP-binding: Repeat.
CC NP_BIND 231
CC NP_BIND 508 515 ATP (POTENTIAL).
CC FT BIND 508 515 ATP (POTENTIAL).
CC SEQUENCE 745 AA; 83138 MW; B195B2044FCF1F2 CRC64;

Query Match
Best Local Similarity 21.3%; Score 140.5; DB 1; Length 745;
Matches 93; Conservative 70; Mismatches 148; Indels 125; Gaps 21;

64 RAVLAGPATKTLALGIAOELGSVYPCPMWSEYV-----SEVKTEVLMENFR 118
225 KGVILXPGPGTKTLARAVANESGAN--FLSINGPELMSKYVGSOKLRELPSKAET 282

119 AIGLRKENKEYEVEVELSPKPEAESTTGVAKSISHVLSLTKVGTOKLDDSIYD 178
283 APSI-----FIDEIDSIAPKREE-----VOGEVRRVAQLLT 316

179 AL--IKK--VAAGDVITYEANSKAVKRGCDSPATEYDLAEYVPIPKGEVHKKE 233
317 LMGMKRGHVIVIGATNTRIDAIDPALRPRG-----FDREIEIGVDRNG--RKE 365

234 IVODVTLHDLDAANAPOG-----GODILSLMGQ----- 262
366 ILM---IH---TRNPLDMSEEEKKLEEMADTYTGVGADLALVRESAMNLRRL 418

263 ---MMKPKTEITEK-----ROEINKVYNYIDEGIAELVGVLFIDEVHMDIEC 311
419 PEIDLDKPIPTLELKMVYTEDFNALKSIEPSSIREYVVE-VNVYHMDIDIGLEVK- 476

312 FSYLNRALESP-LSPIYIILATNNGICNVGDTMSPHGIPVLDLRLVIRRET-----Y 365
477 -REIKETVELPLKPDVE--KRUGIRSPKGFELYGPPGKTKLLKAVATEESNANPISIK 533

366 GPFEMQIL-----AIRAQVEIDMDESLAYLGEIGQOTSLRHAQIOLISPAVSYSKING 420
534 GPEVLKMWGESEKAHEIRKAKOVAPALVFLDEIDS-----IAPRGTTSDSG 583

421 -REKICRADLEEVSG 435
584 VTERIVNQLTSLDGI 599

RESULT 8
HSDU_BACHD STANDARD; PRT; 464 AA.

AC G9KA27;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ATP-DEPENDENT HSL PROTEASE ATP-BINDING SUBUNIT HSLU.
 GN HSLU OR CLP1 OR BH2463.
 OS Bacillus halodurans.
 OC Bacillus/Staphylococcus group; Bacillus.
 NCBI_TaxID=86665;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kohara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 CC -1- FUNCTION: CHAPERONE SUBUNIT OF A PROTEASOME-LIKE DEGRADATION
 CC COMPLEX (BY SIMILARITY).
 CC -1- SUBUNIT: INTERACTS WITH HSLV (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOSOL (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CLPX CHAPERONE FAMILY. HSLU SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: AP001515; BAB06182.1;
 KW Chaperone; ATP-binding.
 FT NP_BIND
 SQ SEQUENCE 464 AA; 52634 MW; 6575FED65823191 CRC64;

Query Match 6.1%; Score 139; DB 1; Length 464;
 Best Local Similarity 20.6%; Pred. No. 0.11; Mismatches 122; Indels 126; Gaps 23;

Matches 84; Conservative 76; Mismatches 122; Indels 126; Gaps 23;
 QY 40 VQQAAREAGLAV-----DMIRQKMGAGRAVLAPATGKTALAGIAOELGS 89
 DB 19 VQDEGAKRSVAIALNNRRRFOLEKLD-ETPKNITMIGTGVKTEIARLAKLVGA 77
 QY 90 KYRPPCM-----VGSEVYS--SEVKTEVLMENFRALIGLRKEKVEYEGEYVE 137
 DB 78 --PFVKVETATKTEVGYGRVDSMINDVLTSVRLVKEEKKMGVKKQAEQANORIVEL 135
 QY 138 LSPFEASTGGYAKSISHVIISLTKYKTKOLKLD-----SSIYDAL-----IKKVVAN 187
 DB 136 LVP--AKKKTAYKN-----PLEMLFGOOOEWDTHQETSADODLTERRRMSQDLAL 186
 QY 188 GDVIYEANSQAVKRGVCSFATFEDYLEAEVYPIPKGVHKKELVODVTLHIDDAAN 247
 DB 187 GELE-----DHYVY--VEVEEOTP-----OFFDMLQGS 212
 QY 248 AOPGQDILSLMGOM-----KPKKTEITTEKL-----QETKNVNVYIDEGLAELVP-- 296
 DB 213 GMEQMGMMKMGMMKPKKRRKRLVREARVLTVEEOKLID--MEVQOEAVSAE 270
 QY 297 --GVLTIDEVHMT-----DIECSYLNRALESPLSYIVLATNRGICNVRGDTMSP 346
 DB 271 OLGIIVFIDEIDKTAGGOSADVS-----REGVORDILPI-----VEGSTVYTK 314
 QY 347 HGIVPDLRLVITRTETV---GPEMIQIIL-----AIRAOVEIDMDE 387
 DB 315 YG--PVS--TDHMLFAGAFHVAKPSDLIPELGRRPIVELSGLNLTVD 360

RESULT 9
 ID AFG2_YEAST
 AC P32794;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE AFG2 PROTEIN.
 GN AFG2 OR YLR397C OR L8084.16.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94152174; PubMed=8109176;
 RA Thorsness P.E., White K.H., Ong W.-C.;
 RT "AFG2, an essential gene in yeast, encodes a new member of the
 RT Sec18p, Psp1p, Cdc48p, Tbp-1 family of putative ATPases."
 RL Yeast 9:1267-1271(1993).
 CC [2]
 CC SQUENCE FROM N.A.
 CC STRAIN=S288C / AB972;
 CC Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 CC Favell A., Fulton L., Gallung S., Greco T., Kirsten J.,
 CC Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 CC Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 CC Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
 CC Riffen L., Riles L., Tatch A., Trevisan E., Vignati D.,
 CC Wilcox L., Wolfdman P., Vaudin M., Wilson R., Waterston R.;
 CC Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: NOT YET KNOWN. ESSENTIAL FOR VIABILITY.
 CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: L14615; AAC37367.1;
 DR EMBL: U19729; AAB82355.1;
 DR PIR: S39110; S39110.
 DR SCD: S0004389; AFG2.
 DR InterPro: IPR001939;
 DR Pfam: PF00004; AAA; 2.
 DR PROSITE: PS00674; AAA; 2.
 DR ATP-binding; Repeat.
 FT NP_BIND
 FT NP_BIND
 FT NP_BIND
 SQ SEQUENCE 780 AA; 84747 MW; 75094DFA30401D4E CRC64;

Query Match 6.1%; Score 139; DB 1; Length 780;
 Best Local Similarity 24.3%; Pred. No. 0.33;
 Matches 62; Conservative 44; Mismatches 89; Indels 60; Gaps 12;

QY 62 AGRAVLGAPPAKGTALAGIOELGSKVPPGMGSEVYSSEVKTE-VLMENFRRAI 120
 DB 549 APKGVLLYGPGRGSKTLTKALATSEG--INFLAYKGPETFKYGESERAIREIFRA- 605
 QY 121 GLAIKEKVEYEGEVELSPFEASTGGYAKSISHVIIS-LTKYKTKOLKLDSSIYDA 179
 DB 606 --RSAAPSIIFEDIDALSPDRDGSST--SAANHVLISLNLNEDGYEELK----- 651
 QY 180 LIKEKVAVDVYIEAN-----SGAVKVRGDSFATFEDYLEAEVYPIPKGVHKKKEI 234
 DB 652 -----GVIVYATNRPDEIDALLRPRGLD-----RHIVGPPVNNARLEI 692
 QY 235 VDDVT-----LHDLDAANAOPGQDILSLMGOMMKPKKTEITTEKLQOEINKVY 283

```

QY      64  RAVILAGPAPGATATLAIAGIELSGKVPFCPMWGSE-----YSEVAKTEYLEMNT 116
      : : || || || || || || : : || : : || : : || : : || : : || : : ||
Db     220  KGILLVGPPTGKTLAKAIANE--ADVFPSVASEPEFMELIGCAARVR-----DLF 271
      : : || || || || || || : : || : : || : : || : : || : : || : : ||
QY     117  RRAIGLRKEKE--VEGEVTELSDEAEASTTG--YAKSIHWIISLKTIVGTQOLK 171
      : : || || || || || || : : || : : || : : || : : || : : || : : ||

```

CC -----
DR EMBL; X56956; CAA40276.1; -.
DR EMBL; Z74174; CAA98694.1; -.
DR PIR; A39977; A39977.
DR HSSP; P02998; 1AH9.
DR SGD; S0002284; CDC48.
DR InterPro; IPR001939; -.
DR Pfam; PF00004; AAA; 2.
DR PROSITE; PS00674; AAA; 2.
KW Cell cycle; ATP-binding; Repeat.
FT NP_BIND 255 262 ATP (POTENTIAL).
FT NP_BIND 528 535 ATP (POTENTIAL).
SQ SEQUENCE 835 AA; 91995 MW; 02ADDB9A227614D8 CRC64;

RESULT	12
FTSH_PORPU	
ID	FTSH_PORPU
AC	P51327;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	15-JUL-1996 (Rel. 36, Last annotation update)
DE	CELL DIVISION PROTEIN FTSH HOMOLOG (EC 3.4.24.-).
GN	FTSH OR YCE25.
OS	Porphyra purpurea.
OG	Chloroplast.
OC	Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX	NCBI_Taxid=2787;
RP	SEQUENCE FROM N.A.
RC	STRAIN=AVONPORT.
RA	Reich M.E., Munholland J.;
RT	"Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.";
RT	Plant Mol. Biol. Rep. 13:333-335(1995).
RL	-I- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE
CC	(BY SIMILARITY).
CC	-I- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
CC	-I- SUBCELLULAR LOCATION: CHLOROPLAST; INTEGRAL MEMBRANE PROTEIN
CC	(POTENTIAL).
CC	-I- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4.1 (ZINC METALLOPROTEASE).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC	or send an email to license@isb-slb.ch).
CC	-----
DR	EMBL: U38804; AAC08213.1; -.
DR	MEROPS: M41.005; -.
DR	InterPro: IPR000130; -.
DR	InterPro: IPR000642; -.
DR	InterPro: IPR01939; -.
DR	Pfam: PF00004; AAA; 1.
DR	Pfam: PF01434; Peptidase_M41; 1.
DR	PROSITE: PS00674; AAA; 1.
DR	PROSITE: PS00142; ZINC_PROTEASE; 1.
KW	Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;
KW	Zinc; Chloroplast.
FT	TRANSMEM 8
FT	FTS 28
FT	TRANSMEM 119
FT	NP_BIND 213
FT	ATP (POTENTIAL).
FT	ATP (POTENTIAL).
FT	ZINC (CATALYTIC) (BY SIMILARITY).

QY	39	FWGGAAREAGAEAGVADVMIRQKMGAGRAVLAGLPANGKATLALGTADELGSVPPCPWVG	9
Db	191	FLKQPESTFAVAKKIP-----KGVLLGPPGTGKTLTALAALGANG--VPPFSISG	2
QY	99	SE-----VYSEYKKEEVLMEFNFRRAIGLRIKENKE--VEEGEYTELSPDEASTGG	1
Db	240	SEFEYMEVGVGASPRV-----DLFKKA-----KDNAPCIFIDELIDAVGRQGTGVGG	2
QY	150	---YAKSISHVLIISLTKYKTKRQLKLDSSIYDLALIKEVAVGDVIYIAN-----SGAVR	20
Db	289	NDREEDTLLQOLLTEMDGEGN-----TGVIYAATNRADILDSALL	3
QY	202	RVRGDSFAFTEYDLEAEYVPIPK-----EVH--KKETIVDQVTLHDIDANAPQGGOD	2
Db	330	RPRG-----FDROYSDVDFEGRRLALEVYAKKMKESKVSLETI--ARRPFGSGAD	3
QY	256	IILSMQO-----MMKPRKTEITEKLRQEIKNVNRVIDEGIAELPGVLFIDEVHMLD---	30
Db	382	LANLEMAAILTARRRRKSAMTMS-----EIDNSIDRYVAG--LEGTLIDYNS	42
QY	309	--IECPSYNNRL-----ESPLSPIVILATNRGICANVKRGTDMTSPHGIPVDLLDRVI	35
Db	428	KRLIAYHEYGAAIIGLSLEHHDPQKTLIP--RG--GARGLTWTFPSD-----DQSLI	47
QY	360	IRTEFYGEPTMIOIILAIRAOVEEIDDEE-----SLAYGELGGQOTSLRHAIIQLISPA	41
Db	478	SRSQIL--ARIVGALGGRAADEIIPGDAEVTGASNDLQOYVTSNAROMVTRFQMSKIGPL	53
QY	413	SVYSKTN---GR-----EKCKRADLEEVSLYLDAK	440
Db	536	SLESQSDPFLRGMGSGSEYSDENVATNIDKQVREIVSECYEAK	580

RESULT 13

HSLU_THEME

ID

HSLU_THEME

STANDARD;

PRT;

463 AA.

AC

O9WY22;

DT

01-OCT-2000 (Rel. 40, Created)

DT

01-OCT-2000 (Rel. 40, Last sequence update)

DT

01-OCT-2000 (Rel. 40, Last annotation update)

DE

ATP-DEPENDENT HSL PROTEASE ATP-BINDING SUBUNIT HSLU.

GN

HSLU OR TM0522.

OS

Thermotoga maritima.

OC

Bacteria; Thermotogales; Thermotoga.

OX

NCBI_Taxid=2336;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN=MSB8 / DSM 3109;

RX

MEDLINE=99287316; PubMed=10360571;

RA

Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,

RA

Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,

RA

McDonald L., Utebäck T.R., Malek J.A., Linher K.D., Garrett M.M.,

RA

Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,

RA

Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,

RA

Salzberg S.L., Smith H.O., Venter J.C., Frazer C.M.;

RT

"Evidence for lateral gene transfer between Archaea and Bacteria from

RL

genome sequence of Thermotoga maritima."

RN

Nature 399:323-329(1999).

CC

-1- FUNCTION: CHAPERONE SUBUNIT OF A PROTEASOME-LIKE DEGRADATION

CC

COMPLEX (BY SIMILARITY).

CC

-1- SUBUNIT: INTERACTS WITH HSLV (BY SIMILARITY).

CC

-1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).

CC

-1- SIMILARITY: BELONGS TO THE CLPX CHAPERONE FAMILY. HSLU SUBFAMILY.

OC Caenorhynchus: Metazoa: Nematoda: Chromadorea; Rhabditida: Rhabditioidea;
OC Rhabditidae: Metazoa: Nematoda: Chromadorea; Rhabditidae: Rhabditioidea;
OC NCB1_TextID=6239; Rhabditidae: Rhabditioidea;
RM [1]
RC SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;

RA Oliver S., Coenen S., Skelton S., Squares S., Sgares R., Sulston J.E.,
RA Rutter S., Seeger K., Skelton S., Squares S., Sgares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of *Mycobacterium tuberculosis* from the
RT complete genome sequence.";
RT Nature 393:537-544(1998).
RL

Wed Nov 14 08:34:18 2001

us-09-589-510-4.std.rsp

Page 11

OM of: US-09-589-510-4 to: EST: * out_format: pfs
Date: Nov 13, 2001 7:26 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame-pzn.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09589510/runat_13112001_065633_13437/app_query.fasta_1.518
-DB=EST -OPMT=fastap -SUFFIX=std.rst -GAPOP=12.000 -GAPEXT=4.500
-MIMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEX=7.000 -STRAP=1 -MAPRFX=biolum62 -TRANS=human40.cdl
-LIST=45 -DOCLIGN=200 -THR_SCORE=pet -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pis -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09589510_@CGN1_1_3444 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-589-510-4
Query length: 435

Database sequences: 10228115
Database length: 431459454
Search time (sec): 1407.780000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Mus musculus	10 days
gb_hic:AK012170	+ 1640.00	3278.59	1.7e-173	1595	AK012170 Mus musculus	10 days
gb_est27:AL131743	+ 1601.00	3201.20	3.4e-169	1427	AL131743 B927 097 Trypanosoma	
gb_est29:AL155383	+ 1218.00	2437.33	2.3e-126	994	AL155383 AL155383 LTI_NFL006.P1	
gb_est29:AL155383	+ 1181.00	2356.57	3.0e-122	919	AL155383 AL155383 LTI_NFL006.P1	
gb_est169:BE035216	+ 1143.00	2280.05	7.0e-118	1068	BE035216 M001E02 MO Mesembryant	
gb_est29:AL155383	+ 1118.00	2207.12	3.4e-115	962	AL155383 AL155383 LTI_NFL012.TC1	
gb_est29:AL155383	+ 1106.00	2207.61	7.7e-114	897	AL155383 AL155383 LTI_NFL013.TB1	
gb_est29:AL155383	+ 1100.00	2195.61	3.5e-113	873	AL155383 AL155383 LTI_NFL013.TB1	
gb_est29:AL155383	+ 1077.00	2148.88	1.4e-110	899	AL155383 AL155383 LTI_NFL006.P1	
gb_est28:AL154734	+ 1056.00	2106.49	3.3e-108	899	AL154734 AL154734 LTI_NFL006.P1	
gb_est100:BE937396	+ 1013.00	2020.13	2.1e-103	866	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 1006.00	2006.03	1.3e-102	939	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 1004.50	2001.09	2.4e-102	1012	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 995.00	1987.57	1.4e-101	632	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 971.50	1975.80	6.2e-99	935	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 972.00	1941.25	6.2e-99	626	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 963.00	1922.18	6.0e-98	840	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 955.50	1904.41	5.9e-97	875	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 953.50	1899.98	1.0e-96	868	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 950.50	1891.45	3.1e-96	1066	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 947.00	1889.47	4.2e-96	698	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 946.00	1889.07	4.2e-96	610	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 942.00	1878.01	1.7e-95	896	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 928.00	1848.12	8.0e-94	757	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 927.00	1848.12	8.0e-94	757	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 924.00	1843.64	1.4e-93	664	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 924.00	1840.13	2.2e-93	885	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 916.00	1824.16	1.7e-92	695	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 909.00	1812.81	7.4e-92	728	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 908.00	1810.23	1.0e-91	1000	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 903.50	1797.34	5.4e-91	576	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 893.00	1802.93	2.6e-91	1005	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 893.50	1801.95	4.3e-90	1005	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 889.50	1787.81	5.8e-89	1116	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 888.50	1765.74	3.1e-89	736	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 880.00	1756.97	9.5e-88	555	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 873.00	1738.58	1.0e-87	758	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 872.00	1736.70	1.3e-87	782	BE937396 602439206F1 NIH_MGC_48	
gb_est100:BE937396	+ 860.00	1710.52	3.7e-86	920	BE937396 602439206F1 NIH_MGC_48	

gb_est26:AI878120 + 859.00 1713.51 2.5e-86 606 I AI878120 fc57402.y1 Zebrafis
gb_est192:BF794804 + 836.00 1703.81 8.7e-86 821 I BF794804 602565650F1 NIH_MGC
gb_est179:BE795532 + 854.00 1700.01 1.4e-85 805 I BE795532 601592516F1 NIH_MGC
gb_est80:BE903175 + 850.00 1691.10 4.5e-85 863 I BE903175 601673419F1 NIH_MGC

seq_name: gb_hic:AK012170

seq_documentation_block: 1595 bp mRNA HTC 08-FEB-2001
LOCUS AK012170
DEFINITION Mus musculus 10 days embryo cDNA, RIKEN full-length enriched library, clone:2610528G06, full insert sequence.
ACCESSION AK012170
VERSION AK012170.1 GI:12848750
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA, clone:12610528G06.
ORGANISM Mus musculus
Mammalia: Eutheria: Chordata: Craniata: Vertebrata: Euteleostomi: Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.

REFERENCE 1 (sites)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Methods Enzymol. 303, 19-44 (1999)
2 (sites)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, H., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (11), 1571-1571 (2000)
3 (sites)
Shibata, K., Itoh, H., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Aizawa, K., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Okazaki, Y., Muramatsu, M., Inoue, Y., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system -384-Format
sequencing pipeline with 384 multipipillary sequencer
Genome Res. 10 (11), 1571-1571 (2000)
4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 1595)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

217 aglGluThrValProIleProIleProIleProIleGlyGluVal.HisLysLysLysGlu 233
 704 TGAAGACTACTCCCTTTGGCCAAAG...GAACTTGCACCAAGAAGAGAA 750
 234 TLeValGlnsPValTrhLeuHisAspLeuAspAlaAlaAsnAlaGlnr 250
 751 ATCATACAGGATGTGACCTTGACATGACCTGGATGT.GCTAATGGCCGCC 799
 250 ogInGlyGlnAspTrIleLeuSerLeuMetGlyGlnMetMetLysProA 267
 800 TCAGGGGGGCAAGATATTTCTGTATGAATGGGCGCAATGATGAACCAA 849
 267 rglYsThrGluIleTrhGluLysLeuArgGlnGluIleAsnLysVal 283
 850 AAAAGACAGAGATCACAGATAACTTCGAGGGAGATCAACAAGTGGTG 899
 284 AsnArGTrIleAspGluGlyIleAlaGluLeuValProGlyValLeuph 300
 900 AACAAATACATTGACCAAGGTGGCCGAGCTGTCCCTGGACTGCTTTT 949
 300 eIlleAspGluValHisMetLeuAspTrIleGlyCysPheSerTrIleAsnA 317
 950 TGTGATGTAGGTCCACATGCTGGATATCAAGTCTTACTACTCCACACC 999
 317 rGAlaLeuGlnSerProLeuSerProIleValIleLeuAlaTrhAsnAr 333
 1000 GAGCCTCGAGTCTCCATCGCCCCCATGTGTATCTTTCATCCAAACCA 1049
 334 GlyTrIleCysAsnValArgGlyThr...AspMetTrhSerProHisGlyTrI 349
 1050 GGCAACTGTGTCTACAGGGGCACTAGAGCAACACTTCTCCACAGGCAT 1099
 349 eProValAspLeuAspArgLeuValIleIleArgTrhGluTrhArg 366
 1100 CCGCTTGGATCTGTAGACAGGGTGATATCATCAGAACCATGCTATACA 1149
 366 LyrProTrhGluMetIleGlnIleLeuAlaIleArgAlaGlnValGln 382
 1150 CGCCACAGAGATGAAAGCAGATCATTAAGATCCGAGCCCAAGCAAGGC 1199
 383 IleAspMetAspGlnLysSerLeuAlaTrIleGluGlyGlnIleGlyGln 399
 1200 ATCAACATCACAGAGAGGCCCTGACACCACTGGGGAGATTGGCACCAA 1249
 399 nTrhSerLeuArgHisAlaIleGlnLeuIleSerProAlaSerValVal 416
 1250 GACCACCTCTGAGTATTGGTGACGTGCTGACCCAGCCAACTGCTGCG 1299
 416 eTrpYsTrhAsnGlyArgGlnLysTrIleCysLysAlaAspLeuGluVal 432
 1300 CMAAGATCAATGGCAGAGACAGCATTTAGAAGAGCAAGTGAAGATC 1349
 433 SerGlyLeuTrIleAspAlaLysSerSerAlaArgLeuGlnGln 449
 1350 AGGAGCTCTTCTATGACGCCAAGTCTCTGCCAAGATTTTGGCTGACCA 1399
 449 ngInGluArgTrIle 454
 1400 GCAGGACAGATCATG 1415
 seq_name: gb_est27:AU131743
 seq_documentation_block: 1427 bp mRNA EST 02-MAR-2001
 LOCUS AU131743
 DEFINITION B927 097 Trypanosoma brucei strain 927 ESTs Trypanosoma brucei CDNA
 ACCESSION AU131743
 VERSION AU131743.1 GI:4049548
 KEYWORDS
 EST.
 Trypanosoma brucei.
 Trypanosoma brucei
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.

LOCUS AL553383 994 bp mRNA EST 16-FEB-2001
 DEFINITION AL553383 LTI_FL006_PL2 Homo sapiens cDNA clone CS0D1075YM04 5
 prime, mRNA sequence.
 ACCESSION AL553383
 VERSION AL553383.1 GI:12893166
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 994)
 AUTHORS Li W.B., Gruber,C., Jessee,J. and Polayes,D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 Source
 Location/Qualifiers
 1..994
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0D1075YM04"
 /clone_lib="LTI_FL006_PL2"
 /issue_type="placenta"
 /note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : filiang@life-tech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 281 a 207 c 307 g 196 t 3 others
 ORIGIN

alignment_scores:
 Quality: 1218.00 Length: 312
 Ratio: 4.157 Gaps: 1
 Percent Similarity: 93.910 Percent Identity: 78.526

alignment_block:

US-09-589-510-4 x AL553383 ..

Align seg 1/1 to: AL553383 from: 1 to: 994

```

1 MetArgIleGluGluValGlnSerThrSerLysGlnArgLeuAlaThr 17
|||||
63 ATGAAGATTGAGGAGGTGAAGACACTACGACGACGCGCTCGCTC 112
|||||
17 rHisThrHisIleLysGlyLeuAspAlaAsnGlyMetAlaIleA 34
|||||
113 CCACAGCCACGTGAAGGCTGGGCTGACGAGAGCGGCTGGCAACG 162
|||||
34 LalaAlaAlaGlyPheValGlyGlnAlaAlaAlaArgGluAlaGly 50
|||||
163 AGGGGGCGCTCAGGCTGTGGGCGAGAGACGCGGAGGCGCTATGGC 212
|||||
51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyValAlaValle 67
|||||
213 GTCATAGTGAATTAATCAAAAGCAGAAATGCTGGAAGAGCTGTCTT 262
|||||
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaGlyIleA 84
|||||
263 GTTGGCGGACCTCTGGAACCTGGCAAGACACTGCTGCGCTATGTTG 312
|||||
84 laGlnLysGlySerLysValProPheCysProMetValGlySerGlu 100
|||||
313 CTCAGGAGCTGGTGTAGTAAGTCCCTTCTGCCCAATGTTGGGAGTAA 362

```

```

101 ValTrpSerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
|||||
363 GTTACTCAACTGAGATCAGAAAGACAGAGGTGCTGATGAGAACTCCG 412
|||||
117 gArgAlaIleGlyLeuArgIleLysGlyLysAsnGlyValLysGlyGlyG 134
|||||
413 CAGGGCCATTGGGCTGGCAATTAAGGAGCAAGGAAGTTTGTGAAGGTG 462
|||||
134 LuValThrGluLeuSerProGluGluAlaGlySerThrThrGlyGlyTr 150
|||||
463 AAGCACAGAGCTAATCCGCTGTGACAGACAGAAATCCATGGAGAGATAT 512
|||||
151 AlAlaYserIleSerHisValIleIleSerLeuLysThrValLysGlyTh 167
|||||
513 GGCAAAACCATTTAGCCATGTGATCATAGAGCACTAAACAGCCAAAGAAC 562
|||||
167 rLysGlnLeuLysLeuAspSerSerIleTrpAspAlaLeuIleLysGlu 184
|||||
563 CAACAGCTTGAACCTGACCCCGACATTTTGAAGTTTGCAGAAAGAGC 612
|||||
184 ysValAlaValAlaGlyAspValIleTrpIleGluAlaAsnSerGlyAlaVal 200
|||||
613 GAATAGAACCTGAGATGTGATTTACATTGAAGCCAAAGTGGGCGCTG 662
|||||
201 LysArgValGlyArgCysAspSerPheAlaThrGluTrpAspLeuGluAl 217
|||||
663 AAGAGCGAGGCGAGGTGTGATCTATGCTACAGAAATTCAGACTTGAAGC 712
|||||
217 agLysGluTrpValProIleProLysGlyGluValHisLysLysGlyLys 234
|||||
713 TGAAGATGTGTCCTGCTGCCAAAGGGAGTGTGCACAAAGAAAGAAA 762
|||||
234 lvaValGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnPro 250
|||||
763 TCATCCAGATGTGACCTTGATGATGCTGATGTGCTAATGCGCGGCC 812
|||||
251 Gln.GlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLysPro 267
|||||
813 CAGGGGGGACAGATATCTCTGCTCATGATGGCGCACTATGAAGCCAA 862
|||||
267 rGlyThrGluIleThrGluLysLeuArgGlnLysIleAsnLysValVal 283
|||||
863 AGAAGACAGAAATCACAGCAAACTTCAGAGGAGATTAAT.AAGTGTGTG 911
|||||
284 AsnArgTrpIleAspGluGlyIleAlaGluLeuValProGlyValLeuPh 300
|||||
912 AACAGATGATGACACACAGGATGTGAGCTGCTGCCGCTGTCTGTT 961
|||||
300 eLlaAspGluValHisMetLeuAspIleGlyCys 311
|||||
962 TGTGATGAGGTGTC.ATGCTGGACATTTGAGTGC 994
|||||
seq_name: gb_est29.AL542244
seq_documentation_block:
LOCUS AL542244 919 bp mRNA EST 16-FEB-2001
DEFINITION AL542244 LTI_FL002_PL1 Homo sapiens cDNA clone CS0D080YR12 5 prime
prime, mRNA sequence.
ACCESSION AL542244
VERSION AL542244.1 GI:12874097
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 919)
AUTHORS Li W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

```

FEATURES
source

Location/Qualifiers
1..919
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE08Y012"
/clone_lib="LTI_FLO02_PL1"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 274 a 188 c 279 g 175 t 3 others
ORIGIN

alignment_scores:
Quality: 1181.00 Length: 301
Ratio: 4.188 Gaps: 1
Percent Similarity: 93.688 Percent Identity: 77.741

alignment_block:
US-09-589-510-4 x AL542244 ..

Align seg 1/1 to: AL542244 from: 1 to: 919

```

1 MetAgtllegluuValuAlnSerThrSerLysLysGlnArgTlleAlaTh 17
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
20 ATGAAGATTGAGAGGTGAGAGACACTACGAAAGCCAGCGATCCCTCC 69
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
17 rHstHhIstleLysGlyLeu.GlyLeuAspIlaAsnIleAlaIle 33
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
70 CCACAGCCAGCGTGAAGAGGCTGGAGCAGAGCGGCTGGCCAG 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
34 AlaLeuAlaIleGlyPheValGlyGlnAlaAlaAlaArgGluAlaIle 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
120 CAGCGCGCTCAGGCGCTGTGGGCGCAGAGCGCGAGAGCGCATGTG 169
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
50 yLeuAlaValAspMetIleArgGlnLysMetAlaGlyArgAlaVal 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
170 CGTCATAGTAGAATTATCAAAACCAAGAAATGGCTGGAGAGCTGTCT 219
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
67 euleuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGly 83
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
220 TGTGTGGCAGGACCTCTCGAAGCTGCAAGACAGTCTGGCTCGGCTAT 269
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
84 AlaGlnIleuGlySerLysValProPheCysProMetValGlySerG 100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
270 GCTCAGAGAGCTGGTAGTAAGTCCCTCTGCCCAATGTGGGGAGTGA 319
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
100 uValLysSerSerGluValLysLysThrGluValLeuMetGluAsnPha 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
320 AGTTTACTCACTGAKATCAAGAAGACAGAGGTCTGATGAGACTTCC 369
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
117 rAArgAlaIleGlyLeuArgTlleLysGluAsnLysGluValTyrGly 133
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
370 GGAGGCC.ATTGGGCTGGCAATAAAGAGCAGCAAGAAATTATGAAGT 418
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
134 GluValThrGluLeuSerProGluGluAlaGluSerThrThrGlyGly 150
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
419 GAAGTACACAGAGTAACTCCGGTGAAGACAGAGATCCCATGGAGGATA 468
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
150 rAlaLysSerTlleSerHisValIleIleSerLeuLysThrValLysGly 167
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
469 TGGCAAAACCATTAGCATGTGATGACAGTCAAAACAGCAAGAGGAA 518
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
167 hLysGlnLeuLysLeuAspSerSerTlleTyrAspAlaLeuIleLysGlu 183
   |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

519 CCAACAGTTGAACCTGAGCCCAACATTTTGAAGTTGCAAGAAAG 568
184 LysValAlaValAlaGlyAspValIleTyrTlleGluAlaAsnSerGlyAla 200
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
569 CCAGTAGAAGACCTGAGATGTGATTTACATGACGACACAGTGGGGCGCT 618
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
200 LysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGlu 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
619 GAGAGAGCGGCGCAGGTGTGATGACTATGCCAGAAATTCACCTTGAG 668
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
217 laGluGluTyrValProIleProLysGlyGluValHisLysLysGly 233
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
669 CCGAAGATGATGTCCTTGGCCAAAGGAGATGTGCACAAAAGAAAGAA 718
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
234 lleValGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGln 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
719 ATCATCAAGATGTGACCTTGATGATCTTGATGTGGCTAATGCCGGCC 768
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
250 oGlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLysPro 267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
769 CCAGGGGGGACAAAGATATCTCTGATGATGGCCAGCTAATGAAGCCAA 818
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
267 rGlyThrGluTlleThrGluLysLeuArgGlnGluIleAsnLysValVal 283
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
819 AGAAGACAGAAATCACAGCAAACTTCAGGGGAGATTATTAAGTGGTG 868
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
284 AsnArgTyrTlleAspGluGlyIleAlaGluLeuValProGlyValLeu 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
869 AACAGTACATGACCAAGGAGCATTTCTGAGCTGTGTCCGGTCTCTGT 918
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
300 e 300
919 t 919

```

seq_name: gb_est69:BE035216

seq_documentation_block:

LOCUS BE035216 1068 bp mRNA EST 07-JUN-2000
MO01E02 MO Mesembryanthemum crystallinum cDNA 5' similar to
DEFINITION ttp-interacting protein (tip49), mRNA sequence.

ACCESSION BE035216
VERSION BE035216.1 GI:8330340

KEYWORDS common ice plant.
SOURCE Mesembryanthemum crystallinum

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.

1 (bases 1 to 1068)
REFERENCE 1 (bases 1 to 1068)
Bohnerl,H.O., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea

,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
Scara,G., Wheeler,M. and Zepeda,G.R.
Functional Genomics of Plant Stress Tolerance

Unpublished (2000)
COMMENT Contact: Michalowski,C.B.

University of Arizona
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: cbm@u.arizona.edu

An open reading frame exists.
Insert Length: 1 Std Error: 0.00.

FEATURES
source
Location/Qualifiers
1..1068
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone_lib="MO"
/tissue_type="apical meristem and leaf primordia"
/dev_stage="5 weeks"
/note="no stress"

BASE COUNT 314 a 224 c 281 g 248 t 1 others
ORIGIN


```

51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaValLeu 67
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
226 GTCAATAGTAGAATTATCAAAAGCAAGAAATGCTGGAAGACTGCTT 275
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
276 GTTGCGAGGACCTCTCTGGAACTGCGCAAGACAGCTGCTGCTGGCTATTG 325
84 lAgInGluLeuGlySerLysValProPheCysProMetValGlySerGlu 100
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
326 CTCAGAGACTGGGTAGTAGTCCCTTCTGCCAATAGTGGGAGAGGAA 375
101 ValTyrSerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
376 GTTACTCACTAGATCAAGACAGACAGAGGTGCTGATGAGAACTTCGG 425
117 gArgAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGluGly 134
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
426 CAGGCC.ATTGGGCTGGCAATAAAGAGACCAAGAAAGTTATGAGGTG 474
134 lValTyrThrGluLeuSerProGluGluAlaGluSerThrThrGlyTyr 150
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
475 AAGTCACAGAGCTAATCCGTCGTGAGACAGAAATCCCATGGAGGATAT 524
151 AlAlaLysSerLysSerHisValIleIleSerLeuLysThrValLysGly 167
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
525 GGCAGAAACCATTAAGCATCTGATCTAGACCTCAAAACAGCCAAAGGAC 574
167 rLysGlnLeuLysLysAspSerSerIleTyrAspAlaLeuIleLysGlu 184
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
575 CAACAGATTGAAACCTGACCCCGCATTTTGCAGATTGAGAAAGAGC 624
184 yValAlaValAlaGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
625 GAATAGAGAGCTGGAGATGTGATTTACATTGAAGCCACAGATGGGCGCTG 674
201 LysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGluAl 217
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
675 AAGAGGAGGCGAGTGTGATCTGATCTGACACAGAAATTCGACCTTGAAG 724
217 agLugluTyrValProIleProLysGlyLysValAlaHisLysLysGluI 234
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
725 TGAAGATGATGTCCTCTGCCAAAGGGATGTGCACAAAAGAAAGAAA 774
234 lValGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnPro 250
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
775 TCATCCCAAGATGTGACCTTGACATGTGATGTGCTAATGCGGCGCC 824
251 GlnGlyLysLysAspIleLeuSerLeuMetGlyGlnMetMetLysProAr 267
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
825 CAGGGGAGACAGATATCTGTCATGATGCGGACCTAATGAGCMAAA 874
267 gLysThrGluIleThrGluLysLeuArgGlnGluIleAsnLysValAla 284
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
875 GAAAGACGAATTCAGACAACTTCGAGGGGAGATTAAAGTGTGTGA 924
284 snArg.TyrIleAspGluGlyIleAlaGluLeuVal 295
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
925 MAAGTACAT...CGACAAGCATGTGCTGAGCTGTGC 957
seq_name: gb_est29:AL559327
seq_documentation_block: 893 bp mRNA EST 16-FEB-2001
LOCUS AL559327 L1L_NFL008_Tc2 Homo sapiens cdna CS0D012Y013 5
DEFINITION prime, mRNA sequence.
ACCESSION AL559327
VERSION AL559327.1 GI:12904718
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE 1 (bases 1 to 893)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..893
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="L1L_NFL008_Tc2"
/sex="male"
/tissue_type="T cells from T cell leukemia"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 265 a 186 c 269 g 172 t 1 others
ORIGIN

alignment_scores:
Quality: 1106.00 Length: 280
Ratio: 4.205 Gaps: 0
Percent Simlarity: 93.929 Percent Identity: 77.500

alignment_block:
US-09-589-510-4 x AL559327
Align seg 1/1 to: AL559327 from: 1 to: 893

1 MetArgIleGluGluValGlnSerThrSerLysLysArgIleAlaThr 17
   ::::::::::::::::::::::::::::::::::::::::::::::
53 ATGAAGATTGAGAGAGGTGGAAGACACTAGAGACGCGATGCGCTC 102
17 rHisThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleA 34
   ::::::::::::::::::::::::::::::::::::::::::::::
103 CCACAGCCACGTGAAGGGCTGGGCGTGGACGAGAGCGCTTGCCCAAGC 152
34 lAlaValAlaGlyPheValGlyGlnAlaAlaAlaArgGluAlaIleGly 50
   ::::::::::::::::::::::::::::::::::::::::::::::
153 AGCGGCGCTCAGGCGCTTGCGCCAGAGACGCGCGAAGGCAATGTGGC 202
51 LeuAlaValAspMetIleArgGlnLysLysMetAlaIleArgAlaValLe 67
   ::::::::::::::::::::::::::::::::::::::::::::::
203 GTCAATAGTAGAATTATCAAAAGCAAGAAATGCTGGAAGAGCTGCTT 252
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
   ::::::::::::::::::::::::::::::::::::::::::::::
253 GTTGCGAGGACCTCTCTGGAACTGCGCAAGACAGCTGCTGCTGCTATTG 302
84 lAgInGluLeuGlySerLysValProPheCysProMetValGlySerGlu 100
   ::::::::::::::::::::::::::::::::::::::::::::::
303 CTCAGAGACTGGGTAGTAGTCCCTTCTGCCAATAGTGGGAGAGTGA 352
101 ValTyrSerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
   ::::::::::::::::::::::::::::::::::::::::::::::
353 GTTACTCACTAGATCAAGACAGACAGAGGTGCTGATGAGAACTTCGG 402
117 gArgAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGluGly 134
   ::::::::::::::::::::::::::::::::::::::::::::::
403 CAGGCC.ATTGGGCTGGCAATAAAGAGACCAAGAAAGTTATGAAAGTG 451

```

134 luValThrGluLeuSerProGluGluAlaGluSerThrThrGlyGlyTyr 150
 452 AAGTCACAGAGCTAACTCCGTGACAGACAGAACTCCATGGAGGATAT 501
 151 AlalysSerIleSerHisValIleIleSerLeuLysThrValGlyGly 167
 502 GGCAAAACCATTAGCCATGATCATAGACTCAAAACAGCAAGGAAC 551
 167 rlyGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGlu 184
 552 CAACAGATTGAAACTGACCCCGCATTTTGAAGTTTGCAGAAAGAGC 601
 184 yValAlaValGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
 602 GATTAAGAGCTGGAGATGATGATTACATTGAAGCCACAGAGTGGCGCTG 651
 201 LysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGluAl 217
 652 AAGAGCAGAGGCGAGGTGATACCTATGCCACAGAAATTCGACCTTGAGC 701
 217 aglGluTyrValProIleProLysGlyGluValHisLysLysLysGlu 234
 702 TGAAGAGTATGTCCTTCCCAAAAGGGGAGTGCACAAAAGAAAGAA 751
 234 lValGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnPro 250
 752 TCATCCAAATGTGACCTTGACCTTGATGATGAGCTAATGCGCGGCC 801
 251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyLysMetLysProArg 267
 802 CAGGGGGGCAAGATATCTGCTCATGATGGCCAGCATTAAGAACCA 851
 267 gLysThrGluIleThrGluLysLeuArgGlnIleAsn 280
 852 GAAGAAAGAAATCACAGCAAACTTCAGGGGAGATTAAAT 891
 seq_name: gb_esc129:AL535830

seq_documentation_block: 877 bp mRNA EST 13-FEB-2001
 LOCUS AL535830 LTI_FL013_Fbrn1 Homo sapiens cDNA clone CSDF013Y110 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AL535830
 VERSION AL535830.1 GI:12799323
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 877)
 AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr.
 FEATURES
 source
 1. 877
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSDF013Y110"
 /dev_stage="pooled tissue from post conception fetuses (20
 week, 24 week and 26 week)"
 /lab_host="DH10b"
 /note="Organ: Fetal brain. Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact: Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville

BASE COUNT 258 a 185 c 262 g 166 t 6 others
 ORIGIN
 'Maryland 20850, USA Fax: (1) 301 610 8371 Email:
 fliang@lifetech.com URL:
 http://fulllength.invitrogen.com"

alignment_scores:
 Quality: 1100.00 Length: 276
 Ratio: 4.264 Gaps: 0
 Percent Similarity: 93.478 Percent Identity: 76.449

alignment_block:
 US-09-589-510-4 x AL535830 ..

Align seg 1/1 to: AL535830 from: 1 to: 877

1 MetArgIleGluGluValGlnSerThrLysLysArgIleAlaThr 17
 44 ATGAAGATTGAGGAGGTGAAGACACTACGAGACGCGCATCGCTC 93
 17 rHisThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIle 34
 94 CCACAGCCACGTGAAGAGGCTGGGCTGCAGAGAGCGGCTTGCCAGC 143
 34 lAlaValAlaGlyPheValGlyGlnAlaAlaAlaArgIleAlaGly 50
 144 AGCGCGCTCAGGCGCTGTGGCCAGACGAGACGCGCAGAGCATGGGC 193
 51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaVal 67
 194 GTCATAGTAGAATTAATCAAAAGCAAGAAATGCTGGAAGATGCTCT 243
 67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaGlyIle 84
 244 GTTGGCAGACCTCCTGGAAGCTGCAAGACAGCTGCTGCTGCTATG 293
 84 lArgIleGluLysSerLysValProPheCysProMetValGlySerGlu 100
 294 CTCAGAGAGCTGGTAAATTAAGCTCCCTTGCCTCAATGGGGAGCTGA 343
 101 ValTyrSerSerGluValLysLysThrGluValLeuMetGluAsnPhe 117
 344 GTTACTCAACTGAGATCAAGAGACAGAGGCTGCTGAGAGACTCCG 393
 117 gArgAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGlyGly 134
 394 CAGGCGCATTTGGCTGGAATTAAGACCAAGAAAGTTATGAAGTG 443
 134 luValThrGluLeuSerProGluGluAlaGluSerThrThrGlyGlyTyr 150
 444 AAGTCACAGAGCTAACTCCGTGACAGACAGAACTCCATGGAGGATAT 493
 151 AlalysSerIleSerHisValIleIleSerLeuLysThrValGlyGly 167
 494 GGCAAAACCATTAGCCATGATCATAGACTCAAAACAGCAAGGAAC 543
 167 rlyGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGlu 184
 544 CAACAGATTGAAACTGACCCCGCATTTTGAAGTTTGCAGAAAGAGC 593
 184 yValAlaValGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
 594 GATTAAGAGCTGGAGATGATGATTACATTGAAGCCACAGAGTGGCGCTG 643
 201 LysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGluAl 217
 644 AAGAGCAGAGGCGAGGTGATACCTATGCCACAGAAATTCGACCTTGAGC 693
 217 aglGluTyrValProIleProLysGlyGluValHisLysLysLysGlu 234
 694 TGAAGAGTATGTCCTTCCCAAAAGGGGAGTGCACAAAAGAAAGAA 743

234 leValGlnaspValThrLeuHisaspLeuaspAlaalaAsnAlaGlnPro 250
|||||
744 CATTCGAAGATGTGACCTTGATGATGCTGATGCTATGCGGCCCC 793
251 GlnGlyGlnaspIleLeuSerLeuMetGlyGlnMetMetLysProAr 267
|||||
794 CAGGGGGGACAGATATCTGCTCATGATGCGCCAGCTAATGAAGCCAA 843
267 gLysThrGlnIleThrGlnLysLeuArg 276
|||||
844 GAGACAGAAATCACAGACAACTTCA 871
seq_name: gb_est197.BG126632
seq_documentation_block: 764 bp mRNA EST 31-JAN-2001
LOCUS BG126632
DEFINITION EST1472278 tomato shoot/meristem Lycopersicon esculentum cDNA clone
CTOF12N8 5' sequence, mRNA sequence.
ACCESSION BG126632
VERSION BG126632.1 GI:12626820
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 764)
AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
Hansen,C., Ronning,C. and Tanksley,S.
TITLE Generation of ESTs from tomato shoot/meristem tissue
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.
FEATURES
source
1..764
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CTOF12N8"
/clone_lib="tomato shoot/meristem"
/tissue_type="shoot/meristem"
/dev_stage="developing shoots from 4-6wks old plants"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI, Site_2:
XhoI; Small expanding leaves from the growing tip were
taken from greenhouse plants (4-6wks old TA496). Tissue
was immediately frozen in liquid nitrogen."
BASE COUNT 243 a 124 c 210 g 187 t
ORIGIN
alignment_scores:
Quality: 1077.00 Length: 246
Ratio: 4.525 Gaps: 0
Percent Similarity: 96.748 Percent Identity: 88.211
alignment_block:
US-09-589-510-4 x BG126632 ..
Align seq 1/1 to: BG126632 from: 1 to: 764
1 MetArgIleGluGluValGlnSerThrSerLysLysGlnArgIleAlaTh 17
|||||
22 ATGAAGATAGAGAGGTACATCAATCAATGAAGATCAGAGAAATTGCTAC 71
|||||
17 rHisThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleA 34
|||||
72 TCATATCATCATTAAGAGCCTTGCTTGAAGCCAAATGAAGAGGACACTTC 121

34 lAlaValAlaGlyPheValGlyGlnAlaAlaAlaArgGluAlaAlaGly 50
|||||
122 CATTCGACACTGGGTTTGTAGCTCAGCGCAGCCAGACAGAACTCGGGG 171
51 lAlaAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaVal 67
|||||
172 CTTGTGTGTGATATGATATACCGCAAAAGAAATGCTGGCGGCTTTACT 221
67 lLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaGlyIleA 84
|||||
222 GCTAGCTGTCGCCCTCGTACGCGGAGACACTCTTGTCTTGGTATAT 271
84 lAlaGlnIleuGlySerLysValProPheCysProMetValGlySerGlu 100
|||||
272 CACAAGACCTTGGAAGCAGAGGTTCCATTGTTCATAGTTGTTGGTCCGA 321
101 ValTyrSerSerGluValLysLysThrGluValLeuMetGluAspPheAr 117
|||||
322 GTGTATTTCATCAGAAAGTGAAGAAACTGAGGCTTAATGAAACCTTCG 371
117 gArGAlaIleGlyLeuArgIleLysGlnLysGluValTyrGluGly 134
|||||
372 GCGAGCTATTGTCTCCGTATGAGAAATAGAGAGTTTATGAAGAG 421
134 lAlaValThrGluLeuSerProGluGluAlaGluSerThrThrGlyLysTyr 150
|||||
422 AGGTGACTGACACTATCTCCAGAAAGCGTCAGAGTGTGACAGTGGATAT 471
151 AlAlaSerLysSerHisValIleIleSerLeuLysThrValLysGlyTh 167
|||||
472 GGTAAAGCATTAAGCCATGTTTAATGGCTTAAACGTCAAGAGTAC 521
167 rLysGlnLeuLysLeuAspSerSerLysThrAspAlaLeuIleLysGlu 184
|||||
522 CAACACGTTGAAGCTTGACCCCAACGATATATGCAATATTAAGAGA 571
184 yValAlaValAlaGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
|||||
572 AGGTACTGCTGTGATGATGATTCATTGATGATCAACAGTGGACAGT 621
201 lysAlaValAlaGlyCysAspSerPheAlaThrGluTyrAspLeuGluAl 217
|||||
622 AAAAGAGTGGCAGAGAGTATGCTTTGGCACAGAAATTTGATCTTGAGA 670
217 aGluGluTyrValProLeuProLysGlyGluValHisLysLysGluI 234
|||||
671 .GAGAGTATGTTCACCTTCTTAAGAGAGGTTCACAGAGAAAGAGA 720
234 lAlaValGlnaspValThrLeuHisaspLeuaspAlaAla 246
|||||
721 TAGTCCAGATGTTACATTCATGATGATCTTGATGCTGCA 758
seq_name: gb_est28:AL514734
seq_documentation_block:
LOCUS AL514734 899 bp mRNA EST 13-FEB-2001
DEFINITION AL514734 LTL.NFL006.PL2 Homo sapiens cDNA clone CL08B013ZH07 5
prime, mRNA sequence.
ACCESSION AL514734
VERSION AL514734.1 GI:12778227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 899)
AUTHORS Li,W.-B., Gruber,C., Jessee,J., and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

Quality:	1077.00	Length:	279
Ratio:	4.191	Gaps:	0
Percent Similarity:	92.115	Percent Identity:	75.986

Align seg 1/1 to: AL514734 from: 1 to: 899

1 MetatgllieglluglluvallglnSerThrserylsySgnIatgllleatTh 17
58 ATGAAGATTGAGGAGGTGAACAGCACTACGAAGACGACGCGCATCGCTC 107
17 rHisThrHisIleLysGlyLeuGlyLeuAspAlaasnGlyMetAlaIleA 34
108 CCACAGCCACCTGAAAGGGCTGGGGCTGTGACGAGARCGCGCTTGGCCAA 157
34 lAlenAlaIaIGlYpHeValIglyInAlaIaAlaIaArgIluAlaIaGly 50
158 AGCGCGCTTCAGGGCTTGTGGGCCAGAGAAAGCGGCARAGCATGTGGC 207
51 lenuAlaValAspMetIleArgGlnLysLysMetAlaClyArgAlaValle 67
208 GTCAATAGTAGAATTAACTCAAAAGCAAGAAATGTGCTGGAAGAGCTGTCTT 257
67 uLeuAlaClyProProAlaThrngLysLysThAlaLeuAlaLendLysIleA 84
258 GTGGCAGAGACTCTCTGGAACTGGCCARACAGCTGTGGCTGTGGCTATTG 307
84 lAgInGluLeuGlySerLysValProPhecysPrometValGlySerglu 100
308 CTCAGGAGACTGGTAGTAAAGTCCCTTCTCCCAATAGTGGGGAGTGA 357
101 ValTyIserSergluValLysLysThGluValLeuMetGluasnPheAr 117
358 GTTAACTCACTGAKATCAARAAACACARAGGTGCTGATGAGAACTTCCG 407
117 gArgAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrgLug 134
408 CAGGCC .ATTGGGCTGCCAATAAAGGAGACCAAGCAAGATTATGAAGTGG 456
134 luValThhGluLeuSerProGluGluAlaGlnSerThrThrhnglyGlyTyr 150
457 AAGTACACGAGCTAACTCCGTGTGAGACAGAGAATCCCATGTGAGAGTAT 506
151 AlaySerIleSerHisValIleIleIleSerLeuLysThrValLysGlyTh 167
507 GCGAAACACATTAGCATGTGATCATAGAGACTCAAAACAGCCAAAGGAAC 556
167 rLysGlnLeuLysLeuaspSerSerIleTyrrAspAlaLeuIleLysGlu 184

```

557 CAACAGGTTGAMACTGGACCCCGACATTTTGAAGTTTGACGAAGAAGC 606
184 ysValAlaValAlGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 2007
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
607 GATTAGACAGCTGGAGATGTGATTTACATTTGAAGCCCAACAGGGGGCCGTG 656
201 LysArgValAlaGlyGlyAspSerPheLeuPheGluTyrAspLeuGluAl 217
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
657 AAGAGCGAGGGCAGGGGTGTATACCTATGCCACAGCATTCGACCTTGAAGC 706
217 aglGluTyrValProIleProLysGlyGlyValAlaHisLysLysLysGlu 234
707 TGAAGAGTATGTGCCCTTGSCAAAAGGGAGTGTGCACAAAAGAAGAA 756
234 LeValGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnPro 250
757 TCATCCCAAGATGTGACCTTGATGACTTGATGTGGCTAAATGCGCGGCC 806
251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLeuLysProAr 267
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
807 CAGGGGGGAGCAAGATATCTGTCCATGTATGGGGCCAGCTAAAGAAGCCAAA 856
267 GlyThrGluIleThrGluLysLeuArgGlnGluIle 279
   ::|||:::||||:::||||| ||||| |||||
857 GAAGACAGATATTCACGACGACAAACCTTCGAGGGGAGACATT 893

```

seq_name: gb_est100:BG397396

```
seq_documentation_block:
```

DEFINITION	602439206F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4565521 5' , mRNA sequence.
------------	--

ACCESSION	BG397396
VERSION	BG397396.1
	GI:13290844

NETWORKS	ESI.
SOURCE	human

ORGANISM	Homo sapiens
Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:	

REFERENCE 1 (bases 1 to 899)

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
unpublished (1955)
COMMENT
Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M Staudt
pb D

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LBNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Console
http://image.ln.gov

File: DECM1202 10M: 1 column
High quality sequence stop: 856.

Source

1. .899

```
/organism="Homo sapiens"  
/db xref="taxon:9606"
```

```
/clone="IMAGE:4565521"
/clone 14b="NTW MCC 48"
```

/tissue_type="primary B-cells from tonsils (cell line)"

```

/ note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;

```

Directionally cloned into EcoRI/XhoI sites using the site₂: EcoRI; cDNA made by 311g8-dr priming.

following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1 kb library constructed by Lina

Hong in the laboratory of Gerald M. Rubin (University of

(Stratagene) and Superscript II RT (Life Technologies).

Note: CHS IS A NIH-MGC LIB

BASE COUNT
ORIGIN

267 a	200 c	247 g	185 t
-------	-------	-------	-------

alignment_scores:

Quality: 1056.00 Length: 285
 Ratio: 4.093 Gaps: 1
 Percent Similarity: 90.526 Percent Identity: 74.035

alignment_block:
 US-09-589-510-4 x BG397396 ..

Align seg 1/1 to: BG397396 from: 1 to: 899

```

93 PheCysPrometValGlySerGluValTyrSerSerGluValLysLysTh 109
|||||.....|.....|.....|.....|.....|.....|.....|.....|
3 TTCTGCCCATGTGTGGGAGTGAAGTTTACTTCACATGAGATCAAGAC 52
109 rGluValLeuMetGluAsnPhenArgAlaIleGlyLeuArgIleLysG 126
|||||.....|.....|.....|.....|.....|.....|.....|.....|
53 AGAGGTGCTGATGAGAACTTCCGAGGCGCATTTGGCTGGGAATAAAG 102
126 LuAsnLysGluValTyrGluGluValIleThrGluLeuSerProGlu 142
|||||.....|.....|.....|.....|.....|.....|.....|.....|
103 AGACCAAGAGTGTATGAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 152
143 AlaGluSerThrThrGlyTyrAlaLysSerIleSerHisValIleI 159
|||||.....|.....|.....|.....|.....|.....|.....|.....|
153 ACAGAGATTCCTCCATGGGAGATATGGCAAAACCATTCACATGATCAT 202
159 eSerLeuLysThrValLysGlyThrLysGluLeuLysLysSerSer 176
|||||.....|.....|.....|.....|.....|.....|.....|.....|
203 AGGACTCAAAACAGCCAAAGCAACCAAGTGAAGTGAAGTGAAGTGAAG 252
176 LeTyrAspAlaLeuIleLysGluLysValAlaValAlaGlyAspVal 192
|||||.....|.....|.....|.....|.....|.....|.....|.....|
253 TTTTGAAG..TTGCAAGAAAGAGCGATGAAGCTGAGAGTGAATTAC 301
193 IleGluAlaAsnSerGlyAlaValLysArgValGlyArgCysAspSer 209
|||||.....|.....|.....|.....|.....|.....|.....|.....|
302 ATTGAAGCCACACAGTGGGCGCTGAGAGCGAGCGAGTGTGATACCTA 351
209 eAlaThrGluTyrAspLeuGluAlaGluGluValValProIleProLys 226
|||||.....|.....|.....|.....|.....|.....|.....|.....|
352 TCCACAGAAATTCGACCTTGAAGCTGAAGAGATGCTCCCTTGCCAAAG 401
226 LysGluValHisLysLysLysGluIleValGlnAspValThrLeuHis 242
|||||.....|.....|.....|.....|.....|.....|.....|.....|
402 GGGATGTCCACAAAAAAGAAATCATCCAAAGATGTGACCTTGCAATG 451
243 LeuAspAlaAlaAsnAlaGlnProGlnGlyGlnAspIleLeuSerLe 259
|||||.....|.....|.....|.....|.....|.....|.....|.....|
452 TTGGATGTGGCTAATGCCGCTGCCAGGGGGGACAAAGATATCTGTCCAT 501
259 uMetGlnMetMetLysProArgLysThrGluIleThrGluLysLeu 276
|||||.....|.....|.....|.....|.....|.....|.....|.....|
502 GATGGCCAGCTAATGAAGCCAAAGAACAGAAATCACAACAACATC 551
276 rGlnGluIleAsnLysValValAsn..ArgTyrIleAspGluGlyIleAl 292
|||||.....|.....|.....|.....|.....|.....|.....|.....|
552 GAGGGAGATTAAATGAAGTGTGAACCAAGTACATCGACCGGGCATTTGC 601
292 aGluLeuValProGlyValLeuPheIleAspGluValHisMetLeuAsp 309
|||||.....|.....|.....|.....|.....|.....|.....|.....|
602 TGAGCTGGTCCCGGCTGTCTGCTTTTGTGATGAGGTCCACATGCTGCACA 651
309 IleGluCysPheSerTyrLeuAsnArgAlaLeuGluSerProLeuSerPro 325
|||||.....|.....|.....|.....|.....|.....|.....|.....|
652 TTGAAGTCTTCACTACCTGCACCGCGCTGAGATCTTATATCGCTGCC 701
326 IleValIleLeuAlaThrAsnArgGlyIleLysAsnValArgGlyThrAs 342
|||||.....|.....|.....|.....|.....|.....|.....|.....|
702 ATCGTCATCTTTGATCCACGAGGCACTTGTGTATCAGAGGACCTGAG 751
342 pMetThrSerProHisGlyIleProValAspLeuAspArgLeuValI 359

```

seq_name: gb_est29:AL551666

```

752 GACATCAATCCCATTCACGATCCTCTTGACCTTCTGACAGCATGATGA 801
359 leIleArgThrGluThrTyrGlyProThrGluMetIleGlnIleLeuAla 375
802 TAATCCGACCAATGCTGTATATCTCACACGCAATGAACAGATCAATTA 851
376 Ile 376
852 ATC 854

```

```

seq_documentation_block:
LOCUS      AL551666      866 bp      mRNA      EST      16-FEB-2001
DEFINITION AL551666 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1063J05 5
prime, mRNA sequence.
ACCESSION  AL551666
VERSION    AL551666.1 GI:12889833
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 866)
            Li W.-B., Gruber C., Jessee J. and Polayes D.
            Full-length cDNA libraries and normalization
            Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES   location/Qualifiers
            source
                1..866
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CS0D1063J05"
                /clone_lib="LTI_NFL006_PL2"
                /rissue_type="placenta"
                /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                was primed with a NotI-oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                vector. Library was normalized. Library was constructed by
                Life Technologies. Contact : Feng Liang Life Technologies,
                a division of Invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                Email : filiang@lifetech.com URL :
                http://fulllength.invitrogen.com"
BASE COUNT  247 a 183 c 266 g 168 t 2 others
ORIGIN

```

alignment_scores:

Quality: 1013.00 Length: 263
 Ratio: 4.118 Gaps: 1
 Percent Similarity: 93.536 Percent Identity: 76.806

alignment_block:
 US-09-589-510-4 x AL551666 ..

Align seg 1/1 to: AL551666 from: 1 to: 866

```

1 MetArgIleGluGluValGlnSerThrSerLysLysGlnArgIleAlaThr 17
|||||.....|.....|.....|.....|.....|.....|.....|.....|
74 ATGAAGATTTGAGAGGTGAAGACACACTACGAAGACCGCATCCCTC 123
17 rHisThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleAla 34
|||||.....|.....|.....|.....|.....|.....|.....|.....|
124 CCACAGCCACGTGAAGGGCTGGGCTGGACGAGACGGCTGGCCAAAC 173
34 laLeuAlaIleGlyPheValGlyGlnAlaAlaIleArgGluAlaIleGly 50
|||||.....|.....|.....|.....|.....|.....|.....|.....|

```

```

174 AGCGGGCTCAGGGCTTTGGGCCAGAGAACGGCGGARGAGCATGTGGC 223
51 LeuAlaValAspMetIleArgIleLysMetAlaGlyArgAlaValIle 67
::: |||:::|||||:::|||||:::|||||:::|||||:::|||||
224 GTCATAGTAAGTAATTAACAAGCAAGAAATGGCTGGAGAGACTGTCTT 273
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIle 84
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
274 GTTGGCAGGACCTCTGGAACTGGCAAGACAGCTGTGGCTGTGCTATTG 323
84 IaGlnGluLeuGlySerLysValProPheCysProMetValGlySerGlu 100
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
324 CTCAGGAGCTGGGTAGTAAGTCCCTCTGCCCAATGGTGGGAGACTGAA 373
101 ValTyrSerSerGluValLysLysThrGluValLeuMetGluAsnPhe 117
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
374 GTTTACTCAACTGAGATCAAGAGACAGAGTGTCTATGGAGAACTCCG 423
117 gATGAlaIleGlyLeuArgIleLysGluAsnLysGluValAlaTyrGlu 134
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
424 CAGGGCCATTGGGCTCGAATTAAGAGACCAAGAAATTATGAAGTG 473
134 LuValThrGluLeuSerProGluGluAlaGluSerThrThrGlyGly 150
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
474 AAGTCACAGAGCTAAGCTCCGTGTGACAGAGAAATCCCATAGGAGG 523
150 rAlaLysSerIleSerHisValIleIleSerLeuLysThrValLysGly 167
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
524 TGGCAAAACCATTAAGCATGTGATCATAGACATCAAAACGCCAAAGAA 573
167 hrLysGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGlu 183
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
574 CCAACAGTTGAAGCTGACCCACCATTTTGAAGTTTGCAGAAAGAG 623
184 LysValAlaValGlyAspValIleTyrIleGluAlaAsnSerGlyAla 200
::: |||::: |||::: |||::: |||::: |||::: |||::: |||
624 CGAGTAAAGAGCTGAGATGTGATTACATGAAGCCAAAGTGGGGCCG 673
200 lLysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGlu 217
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
674 GAAGAGGACGAGCGGTGTATACCTATGACACAGAAATTCGACCTTGA 723
217 IaGlnGluTyrValProIleProLysGlyGluValHisLysLysGlu 233
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
724 CTGAGAGGTATGTCCCTGCCAAAGGGGATGTGCACAAAGAAAGAA 773
234 lIleValGlnAspValThrLeuHisAspLeuAspAlaAlaAsnIle 250
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
774 ATCATCCAGAGATGTGACCTTGACATGTGATGTGGCTAAATGCCGG 823
250 oGlnGlyGlyGlnAspIleLeuSerLeuMetGlyGln 262
| |||||:::|||||:::|||||:::|||||:::|||||
824 CAGGGGAGCAAGATATCTCTGTCATGATGAGGCGAG 859

```

seq_name: gb_est80:BE901271

seq_documentation_block: 939 bp mRNA EST 29-SEP-2000

LOCUS BE901271 939 bp mRNA EST 29-SEP-2000

DEFINITION 601675912F1 NIH_MGC_21 Homo sapiens CDNA clone IMAGE:3958683 5', mRNA sequence.

ACCESSION BE901271

VERSION BE901271.1 GI:10390287

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NIH-MGC http://mgi.nci.nih.gov/

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLNC at: image.llnl.gov

Plate: LICM838 row: 9 column: 04

High quality sequence start: 4

High quality sequence stop: 750.

FEATURES

Location/Qualifiers

1..939

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3958683"

/clone_lib="NIH_MGC_21"

/tissue_type="choriocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: Placenta; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 259 a 208 c 302 g 170 t

ORIGIN

alignment_scores:

Quality: 1006.50 Length: 266

Ratio: 4.075 Gaps: 2

Percent Similarity: 92.857 Percent Identity: 74.436

alignment_block:

US-09-589-510-4 x BE901271 ..

Align seg 1/1 to: BE901271 from: 1 to: 939

```

1 MetaGlyIleGluGluValGlnSerThrSerLysLysGlnArgIleAlaThr 17
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
22 ATGAGATTGAGAGAGAGGAGCACTACGAAAGACGCGCATCGCTC 71
17 rHisThrHisIleLysGlyLeuGlyLeuAspAlaAsnIleMetaAlaIle 34
::: |||::: |||::: |||::: |||::: |||::: |||::: |||
72 CCACAGCCACGTAAGAGGCTGGGGCTGGACGAGAGCGGCTTGGCCAA 121
34 IaLeuAlaAlaGlyPheValGlyGlnAlaAlaAlaArgGluAlaGly 50
|||::: |||::: |||::: |||::: |||::: |||::: |||
122 AGCGCGCTCAGGGCTGTGGGCCAGGAAAGCGCGAGGACATGTGGC 171
51 LeuAlaValAspMetIleArgIleLysLysMetAlaGlyArgAlaValIle 67
::: |||::: |||::: |||::: |||::: |||::: |||::: |||
172 GTCATAGTAAGATTAATCAAAAGCAAGAAATGGCTGGAAGAGCTGTCT 221
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIle 84
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
222 GTTGGCAGGACCTCTCTGGAAGTGGCAAGACAGCTGTGCTGTGCTAT 271
84 IaGlnGluLeuGlySerLysValProPheCysProMetValGlySerGlu 100
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
272 CTCAGGAGCTGGGTAGTAAGTCCCTCTGCCAATGTGGGAGAGTGA 321
101 ValTyrSerSerGluValLysLysThrGluValLeuMetGluAsnPhe 117
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
322 GTTTACTCAACTGAGATCAAGAGACAGAGTGTGCTGATGGAGAACTCC 371
117 gATGAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGluGly 134
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
372 CAGGGCCATTGGGCTCGAATTAAGAGACCAAGAAATTATGAAGTG 421
134 LuValThrGluLeuSerProGluGluAlaGluSerThrThrGlyGly 150
|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

```

422 AAGTCACAGACCTACTCCGTGTGACAGACAGAAATCCCATGGAGATAT 471
151 AAlaysserIIeserHisValIIelIeserleuysThrValIyysgIyTh 167
472 GGCAAAACCATTAGCATGTGATCATAGACTCAAAACAGCAAGAGAGAC 521
167 rlysgIleuIyyspSerSerIeTyAspAlaLeuIleIygsIuL 184
522 CAACAGTGAAGTGAAGGACCCGACATTTTGAAGATTTCAGAAAGAGC 571
184 ySValAlaValAlGlyAspValIIeTyrlIegIuaIaasnserGIyAlaVal 200
572 GAGTAGAGAGCTGGAGATGTGATTACATTGAGCCAAACAGTGGGGCGTG 621
201 LysrIyValAlGlyArgCysAspSerPheAlaThrGIuTyAspIeuGuaI 217
622 AAGAGGACAGGCGAGGTGTGATACCTATGCAACAGAAATTCACCTTGAAGC 671
217 aGIuGIuTyValProIIeProIySGIyGIuValHisIyLysSGIuI 234
672 TGAAGAGATGTCCCTTGGCAAAAGGAGATGTCCCAAAAGAAAGAAA 721
234 lEValAlaAspValIThrleuHisAspLeuAspAlaIaAsnAlaGIuPro 250
722 TCATCCAAAGATGTGACCTGCTGATCTTGATGTGGTAATGCGCGGCC 771
251 GIuGIyGIyGIuAspIIeLeu.SerIeuMeIyGIyGIuMeIeIyS 265
772 AGGGGGGAGCAG...ATATTCTGCCAGATGGGGGAGCAATTATGAG 814
seq_name: gb_est18:BE745673

```

```

seq documentation block:
LOCUS BE745673 1012 bp mRNA EST 15-SEP-2000
DEFINITION 601578913P1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3927650 5',
            mRNA sequence.
ACCESSION BE745673
VERSION BE745673
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1012)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DRP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: L16W757 row: j column: 03
            High quality sequence stop: 8
            Location/Qualifiers
                1..1012
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3927650"
                /clone_lib="NIH_MGC_9"
                /tissue_type="adenocarcinoma cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGCAGCAG(G). Size-selected >500bp for average
                insert size 1.8kb. Library constructed by Ling Hong in
                the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit

```

```

BASE COUNT 314 a 210 c 329 g 159 t
ORIGIN
alignment_scores:
    Quality: 1004.50      Length: 270
    Ratio: 4.067          Gaps: 2
    Percent Similarity: 91.481      Percent Identity: 75.185
alignment_block:
US-09-589-510-4 x BE745673 ..
Align seg 1/1 to: BE745673 from: 1 to: 1012
2 ArgIIeGIuGIuValInsThrSerLysGIuArgIIaIaThrH 18
...
1 AAGACTGAGGAGGTGAAGAGACACTACGAAGACGACACCGCTGCCA 50
18 sThrHisIleIySGIyIeuGIyLeuAspAlaasnGIyMeIaIaIaL 35
51 CAGCCACGTGAAGGGCTGGGCTGGAGAGAGCGCTTGGCCAAAGCAGG 100
35 euaIaIaGIyPheValGIyGlnaIaIaIaArgGIuAlaGIyLeu 51
101 CGGCTCAGGCGCTTGGGCGCAAGAGACGCGAGAGGCAATGTGGCTC 150
52 AlaValAspMetIleArgGlnIySGIyMetAlaGIyArgAlaValLeu 68
151 ATAGTAGAATTATCAAAAGCAAGAAATGGCTGGAAGAGTGTCTGT 200
68 uAlaGIyProProAlaThrGIyLysThrAlaLeuAlaLeuGIyIleAla 85
201 GGAGGACCTCCTGGAACTGGCAAGACACACTGCTGCTGCTATTGCTC 250
85 lInGIuLeuGIySerLysValIProPheCysProMetValGIySerGIuVal 101
251 AGGAGCTGGGTAGTAAGTCCTCCCTTCTGCCAATGGTGGAGATGAGTT 300
102 TyrSerSer.GIuValLysLysThrGIuValIeuMetGIuAsnPheArg 118
301 TACTCAAGCTGAGATCAAGACAGACAGAGTCTGATGAGAACTTCGCA 350
118 rAlaIaIeGIyLeuArgIIeIySGIyAsnIySGIuValIyrcGIuGIy 134
351 GGCCATTGGCTGGCAATTAAGAGACCAAGCAAGTTATTAAGGTGA 400
135 ValThrGIuLeuSerProGIuGIuAlaInsThrThrGIyGIyTYrAl 151
401 GTCACAGAGCTTAAGTCCGTGTGACAGACAGAGATCCCATGGAGATATGG 450
151 alysserIIeserHisValIIelIeserleuysThrValIyysgIyThL 168
451 CAAAACCATTAGCATGTGATCATAGACTCAAAACAGCAAAAGGAGACCA 500
168 ySGIleuIyyspSerSerIeTyAspAlaLeuIleIygsIuIyS 184
501 AACAGTTGAAGTGAAGGACCCGACATTTTGAAGTTTGACAGAAAGAGGA 550
185 ValAlaValAlGlyAspValIIeTyrlIegIuaIaasnserGIyAlaVal 201
551 GTAAGAGCTGAAGATGTGATTACATTGAGCCAAACAGTGGGGCGGTGA 600
201 sArgValAlGlyArgCysAspSerPheAlaThrGIuTyAspIeuGuaIag 218
601 GAGGACAGGCGAGGTGTGATTACCTATGCCAAGAAATTCACCTTGAAGCTG 650
218 lGIuGIuTyValProIIeProIySGIyGIuValHisIyLysSGIuIle 234
651 AAGAGTATGTCCCTTGGCAAAAGGAGATGTGCACAAAAGAAAGAAATC 700
235 ValGIuAspValIThrleuHisAspLeuAspAlaIaAsnAlaGIuProGI 251

```

```

701 ATCCAAGATGTGACTTGCATGACTTGATGTGGCTAATGCGGGCCCA 750
251 nglgylglnaspilleuSerleuMetglglnMetMetLysProArgL 268
||||| :||||| :||||| :||||| :||||| :
751 GGGGACACAGA...TATCTGGCCATGAAGGGGACAGTAAAGCAAGGA 796
268 ysthrglu 270
|||||
797 AGACCGAA 804

seq_name: gb_est71:BE204244

seq_documentation_block:
LOCUS BE204244 632 bp mRNA EST 05-SEP-2000
DEFINITION EST396920 KVO Medicago truncatula cDNA clone pkV0-14N12, mRNA
sequence.
ACCESSION BE204244
VERSION BE204244.1 GI:8747529
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 632)
Vandenbosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L.,
Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E., and
Fraser, C.M.
ESTs from uninoculated seedling roots of Medicago truncatula
Unpublished (1999)
Contact: Vandenbosch K
Department of Biology
Texas A&M University
College Station, TX 77843-3258, USA
Tel: 409 845 7707
Fax: 409 845 2891
Email: katemail.bio.tamu.edu
Texas A&M University name: T264016e
TIGR sequence name: MTGAP78TK
More information is available at:
http://chrysis.tamu.edu/medicago
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
FEATURES
source
1.632
Location/Qualifiers
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pkV0-14N12"
/clone_1b="KVO"
/tissue_type="Seedling roots"
/dev_stage="Immediately prior to inoculation with
Sinorhizobium meliloti (0 hour)"
/lab_host="E.coli strain XL0LR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
stratagene and packaged using gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XL0LR cells."

BASE COUNT 183 a 116 c 160 g 173 t
ORIGIN
alignment_scores:
Quality: 995.00 Length: 210
Ratio: 4.807 Gaps: 0
Percent Similarity: 98.571 Percent Identity: 92.381

alignment_block:
US-09-589-510-4 x BE204244
Align seg 1/1 to: BE204244 from: 1 to: 632

```

```

170 LeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGluLysVal 186
||||| :||||| :||||| :||||| :||||| :
1 CTAAGTTGACCCACAGATATATGATGCTTATCATCAAAAGAGAGC 50
186 aValGlyAspValIleTyrIleGluAlaAsnSerGlyAlaValLysArg 203
||||| :||||| :||||| :||||| :||||| :
51 TGTGGGAGATGTATATATATGAGGCAATAGTGGCGCTGTGAAAAGG 100
203 aLgLYarGcYasPserPheAlaThrGluTyrAspLeuGluAlaGlu 219
||||| :||||| :||||| :||||| :||||| :
101 TTGGCCGAAGTGAATGCTTTGCTACTGAGTGTGACCTTGAAAGCTGAGAG 150
220 TyrValProIleProLysGlyGluValHisLysLysLysGluLeuVal 236
||||| :||||| :||||| :||||| :||||| :
151 TATGTTCCACTTCCCAAGGAGAGGTTCCACAAAGAAAAAGATTTGTCA 200
236 nasPValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnProGlnGly 253
||||| :||||| :||||| :||||| :||||| :
201 GATGTAAACCCCTACATGATCTGGATGCTGCCAACGGCGACTCAAGGGG 250
253 LysGlnAspIleLeuSerLeuMetGlyGlnMetMetLysProArgLysThr 269
||||| :||||| :||||| :||||| :||||| :
251 GGCAGAGTATTTGTCTCTTATGGGCCAGATGAGAACTAGCAAGAGC 300
270 GluIleThrGluLysLeuArgGlnGluIleAsnLysValValAsnArgTy 286
||||| :||||| :||||| :||||| :||||| :
301 GAAATCAGTACAGAGTTGAGACAAAGAAATTAATAGGTTGTCAACGATA 350
286 rIleAspGluGlyIleAlaGluLeuValProGlyValLeuPheIleAspG 303
||||| :||||| :||||| :||||| :||||| :
351 TATCGATGAAGGAGTGTGCAGAACTGTCTCGGGGTTCTCTTATTTAGACG 400
303 LuValHisMetLeuAspIleGluCysPheSerTyrLeuAsnArgAlaLeu 319
||||| :||||| :||||| :||||| :||||| :
401 AGGTGCATATGCTAGATATGAGAGTGTTCCTATCTTAATCGTCTTTA 450
320 GluSerProLeuSerProIleValIleLeuAlaThrAsnArgGlyIleCy 336
||||| :||||| :||||| :||||| :||||| :
451 GAGAGCTCGTGTCTCCCAATAGTAATCTTCGCCCAAAATGAGAGAAATG 500
336 AsnValAlaArgGlyThrAspMetThrSerProHisGlyIleProValAspL 353
||||| :||||| :||||| :||||| :||||| :
501 CACCGTAAGAGGAGCTGATATGACCAAGTCCCTCATGGCATACCTGTTGACC 550
353 euLeuAspArgLeuValIleIleArgThrGluThrTyrGlyProThrGlu 369
||||| :||||| :||||| :||||| :||||| :
551 TATTTGATCGATTGTGATCATTTGCAACACAAACCTAAGGTTCCAGCTGAA 600
370 MetIleGlnIleLeuAlaIleArgAlaGln 379
||||| :||||| :||||| :||||| :||||| :
601 ATATATCAGATTCTAGCTATCCGAGCTCAA 630

```

Wed Nov 14 08:34:19 2001

us-09-589-510-4.std.rst

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2001, 07:57:39 ; Search time 2534.15 Seconds
(without alignments)
11670.318 Million cell updates/sec

Title: US-09-589-510-3
Perfect score: 1912
Sequence: 1 acccagcgctcgcaattt.....aaaaaaaaaaaaaaaa 1912

Scoring table: OLIGO_MUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Genbmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_da1:*
17: em_da2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_r01:*
95: gb_r02:*
96: gb_in4:*
97: gb_pr10:*
98: em_da3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1912	100.0	1912	10	AX077226	AX077226 Sequence
2	1080	56.5	1869	10	AX077232	AX077232 Sequence
3	576	30.1	1845	10	AX077224	AX077224 Sequence
4	94	4.9	1866	10	AX077228	AX077228 Sequence
5	94	4.9	1868	10	AX077230	AX077230 Sequence
6	40	2.1	171809	83	AP003271	AP003271 Oryza sat
7	35	1.8	1173	53	CNS06LVV	AL0404705 T7 end of
8	34	1.8	55	10	129928	129928 Sequence 41


```

Db 901 GAAATCACCAGAAAACACGCCAAGAAATTAATAGGTGTAATATAGATATATCATGAA 960
Qy 961 gaaattcagagccttgcaccgtgtgttttctcatltagatgagccatgttggatc 1020
Db 961 GGAATTCAGAGCTTGTACTCGTGTGTTTGTCTCATGTAGAGGTCACATGTTGGATATC 1020
Qy 1021 gaatttttcttacttaaccgtgcatttggagagccatataccatccgtatattc 1080
Db 1021 GAATGTTTCTTATCTTAACCGTGATGAGAGGCCATTATACCAATCGTATATCTT 1080
Qy 1081 gctcaaaataggaggaataatgtlaagaggaaatgatalgacaagttccacatgtata 1140
Db 1081 GCTACAAATAGGCGAATATGTAATGTAAGAGCACTGATATGACATGACATGCAATGATA 1140
Qy 1141 ccggtgatacttcttagatagctgtgtatattctggcagagagataatgtgcctactgag 1200
Db 1141 CCGGTGATCTTCTTAGATAGCTGTGTATTTGACAGACATATGACATATGACATGATA 1200
Qy 1201 atgatacagatattggtctatccagacaaagtggagagatgtatgataagaaagt 1260
Db 1201 ATGATACAGATATTGGCTATCTCCAGACAACTGAGAGAGATTGATATGATGAGAGAGT 1260
Qy 1261 ctgtctatttaggcagagatcgacagacacatctttagagacatgctatcaattgata 1320
Db 1261 CTGTCTATTATTAGCGAGATCGACAGACAGACATCTTTGAGACATGCTATTCAATTGATA 1320
Qy 1321 tcacctgcagcgtgtgtctcaaaactaatggaagagaagaatctgcaagcgtgactc 1380
Db 1321 TCACCTGCACCGCTGTGTCTCAAAAGCTAATGGAAGAGAAATCTGCAAGCGCTGATCTC 1380
Qy 1381 gaggaaatcagttgggtctatttggatgcaaatccctcgtcgtgcctccagaaacaa 1440
Db 1381 GAGAAATCAGTGGGCTCTATTGTGATGCCAAATCTCGGCTCGCTCCAGAGACAA 1440
Qy 1441 caagaaagatacaccatagatttggatctcgtcgttggaaatctcgaagaagaatgta 1500
Db 1441 CAAGAAAGATCATCATCAGATTTGGATCTCTGCTGTGGAAGTCTCGAAGAGAAATGTA 1500
Qy 1501 gtgtccagcttgaaagtcatctagtcattgactgcttcaacaggtcatagctactg 1560
Db 1501 GTTCCAGCTCGAAAGTCACTAGTGCATGTGCTCTTCAAGGTTCAATAGTCTACTGG 1560
Qy 1561 tcttggagcgagacacatttggagggagacggttgaatttggacgtgctgtgtgta 1620
Db 1561 TCTTGGAGCGACATTTGGGGGGAACGGCTTGAATTTGGAGTGCCTGTGTGTTA 1620
Qy 1621 gtctccaragaagactgtgtccgacatattgctcgttccagcactgtccgtgattaga 1680
Db 1621 GTCTCCARAGAAGACTGTGTCCGGCATATTGCTGCTTCAACGCACTGTGCTGATTAGA 1680
Qy 1681 ttgttccacggttgcagaatgtccgtgtgtgttttacttctgtcatcaggtgtccgaagt 1740
Db 1681 TTGTTCACCGGTGCGAAGTAATTTGCCGTGTGTATCTTGTGCTCAATCGGTGCCGAGT 1740
Qy 1741 ctgctccacggtgtgtattgtggccgaacctactcttgaacacatgataatggatagg 1800
Db 1741 CTGCTCCACGCGTGTGTATTGTGGCCGACCTATCTTTGACCATGGAATAGATGAGG 1800
Qy 1801 attctcaagaatgcaactgtgcatgttattatcttcaaatgtccataaagcataag 1860
Db 1801 ATTCTTACAGAAATGCAACTGTGATGCTTATTATTCTTAATGTCCATTAACATTAAGC 1860
Qy 1861 aaatgttcttcaacmtvtcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912
Db 1861 AAATGTTTCAACAMTMTAAAAAaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912

```

RESULT 2

AX077232 1869 bp DNA PAT 22-FEB-2001
 LOCUS AX077232 1869 bp DNA PAT 22-FEB-2001
 DEFINITION Sequence 9 from Patent WO0105975.
 ACCESSION AX077232

```

VERSION AX077232.1 GI:13121826
KEYWORDS .
SOURCE Zea mays.
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1869)
REFERENCE
  Mahajan, P. B.
  Maize orthologues of bacterial ruvB:cdnas and uses thereof
  Patent: WO 0105975-A 9 25-JAN-2001;
  PIONEER HI-BRED INTERNATIONAL, INC. (US)
FEATURES
  source
    location/Qualifiers
    1..1869
    /organism="Zea mays"
    /db_xref="taxon:4577"
    /note="unnamed protein product"
    /codon_start=1
    /protein_id="CAC32392.1"
    /db_xref="GI:13121827"
  CDS
    64..1380
    BASE COUNT 539 a 373 c 486 g 471 t
    ORIGIN
    337 ggcataccccagagctggcagcaaggtcccttctgtccatgttggatgacagaagt 396
    Db 256 GCGATAGCCCGAGGCTCGCGACGAAGTCCCTTCTGCTATGCTAGATCAAGAGG 315
    Qy 397 tactctcgaggttcaagaagaactgaggtgctgattggaataattccgtagagctatagt 456
    Db 316 TACTCTCGGAGGTCAAGAAACTGAGTGCATGAGAAATTTCCGTAGACTTATAGT 375
    Qy 457 ttgcgtataaaggaacaaagaggttatagaagagaggttaactgaacttcccgaa 516
    Db 376 TTGCGTATTAAGGAACAAAGAGGTTTATGAAGGAGAGGTTACTGAACCTTCCCAGAA 435
    Qy 517 gaggctgaagatcaactgtgtgatatgtaaaagcattagccatgatatcaactgta 576
    Db 436 GAGGCTGAGAGTACAACCTGTGGATATGCAAAAAGCATTAATGCAATGATCAAGCTTA 495
    Qy 577 aagactgttaagggactaagcaactgaagttagatttctcaattatgatatcctgtatc 636
    Db 496 AAGACTGTTAAGGAGCTAAGCAACTGAAGTTAGTTCTCAATTATATGATGCTGTGATC 555
    Qy 637 aaggaagaaggtgagtgagtgatgtatatacatcogaagaacatagtggagcagtgaaa 696
    Db 556 AAGGAAGAAGGTGAGTGTGATGATTTATATACATGCAAGCAAAATAGTGAGACAGTAAA 615
    Qy 697 agagttgtagtattgtattcttctttagaagaatcgactgtgaagcgtgaagaagtatt 756
    Db 616 AGAGTTGTGATGATGATTTCTTTGCTACAGAAATGATCTTGAAGCTGGAAGATGTT 675
    Qy 757 cctatcccccaaggtgaagttcctaagaaaaaagaatgtgtcagagatgtccacttcatt 816
    Db 676 CCAATCCCCCAAGGTGAAGTCCATTAAGAAAAAAGAAATGTGTCAGAGATGTCACTTTCAT 735
    Qy 817 gaccttgatgcagcaaatgtctcagccaaaggtggcgaagatatttgccttatgggc 876
    Db 736 GACCTTGATGCAGCAAAATGCTCAGCCACAAAGGTGGCCAAAGATATTGTTGCTTATGGGC 795

```

QY 877 cagatgataaaccagaaagactgaaatcaccgaaataacgcaagaataaag 936
 |||||||
 Db 796 CAGATGATGAACACGAAAGACTGAAATCACCGAAAACTACGCAAAATAATTAAG 855
 |||||||
 QY 937 gtgttaaatgatatatcgtgaagaatgacagagctgtacactgtgtttgtcatt 996
 |||||||
 Db 856 GTGTAAATGATATATTCGATGAAGAAATTCGAGACTGTACCTGTGTTTGTTCATT 915
 |||||||
 QY 997 gatgaagtcacatgtgtgatalcgaatgttttcttacttaaccgtgcatgtgagagc 1056
 |||||||
 Db 916 GATGAGGTCCACATGTTGGATATCGAATGTTTCTTATCTTAAACCGTCAATGGAGACC 975
 |||||||
 QY 1057 ccaattaccacatgtgtactgttactaacaataggaggaatgttaattgaagaact 1116
 |||||||
 Db 976 CCAATATCCCAATCGATGACTGCTACAAATAGGGGAATGATATGAAGAGAACT 1035
 |||||||
 QY 1117 gatatacaagtcacatggtataccgtgtgactcttactagataggctgtgttatttcg 1176
 |||||||
 Db 1036 GATATGACAGTCCACATGATATACCGGTGATCTTCTGATAGCTGTGATATTTCGG 1095
 |||||||
 QY 1177 acagagacataatgcccactgagatgatacagatatgtgtatccgagcacaagtggag 1236
 |||||||
 Db 1096 ACAGAGACATATGGCCCTACTGAGATGATACAGATATTGGCTATCCGACACAAAGTGAG 1155
 |||||||
 QY 1237 ggaattgatatgataagaagaactctgtcttacttagagagagtcgagacagacatct 1296
 |||||||
 Db 1156 GAGATTGATATGATGATGAGAAAGTCTTGGCTATTATAGGCGAGATCGGACAGACATCT 1215
 |||||||
 QY 1297 ttgagacatgctatcctaattgatalacactgcagcgtgtgtctcaagaactaagtgaaga 1356
 |||||||
 Db 1216 TTGAGACATGCTATTCATATGATATCACTGCGACGCTGTCTCAAAAGACTAATGAGAGA 1275
 |||||||
 QY 1357 gaaagaaatcgcaagagctgtgtctcaggaagtcagtgagctcatttgatgccaaatcc 1416
 |||||||
 Db 1276 GAGAAATTCGCAAGGCTGATCTCGAGAGATTAGTGGGCTTATTGTTGGTCCAAATCC 1335
 |||||||
 QY 1417 tcggtcgtgctgtctcagagagcacaagaagaatacatcactagattgg 1467
 |||||||
 Db 1336 TCGGCTGCGTCTCCAGGAGCAAAAGATACATCATCCTAGATTGG 1386
 |||||||

RESULT 3
 LOCUS AX077224 1845 bp DNA PAT 22-FEB-2001
 DEFINITION Sequence 1 from Patent WO0105975.
 ACCESSION AX077224
 VERSION AX077224.1 GI:13121818
 KEYWORDS
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 1845)
 REFERENCE
 AUTHORS Mahajan, P. B.
 TITLE Maize orthologues of bacterial ruvB:cdnas and uses thereof
 JOURNAL Patent: WO 0105975-A 1 25-JAN-2001;
 PIONEER HI-BRED INTERNATIONAL, INC. (US)

FEATURES
 source
 1. 1845
 /organism="Zea mays"
 /db_xref="taxon:4577"
 85..1452
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC32388.1"
 /db_xref="GI:13121819"

/translation="MREIVQSTSKRORITATHTHIKGLDANGNAIALAAGFVGQSA
 AREAGIAVDMTRKRMAGRAVLVGPATGCTALALQIELGSKFPCPMVSGSEV
 SEEVKTEVIMNFRATIGLRIRKENKVEGVTALSPBEASTGGAKSISHTIIS
 LKTVKCTKOLKIDSSIVDALIKKAVADVITYIENSGAVRVGCDSPATIEYLEAE
 EYVPIPKGVHKKKEIVQDVTYLDHDAANAPQGGODILSLMGQMKRKRKTEITKLR
 QEINRVNRY IDEGIAELVGLFIDEVHNDIECFSTLNRALESPLSIVILATNRG

ICNVRGDMTSPHGIPIVDLRLVIRITETGYCTEMIOILIAQVEDIDIMDESLAY
 LGEIGQOTSLRHAHQILISPAVSYSKINRGRIKADLEBVSGLYLAKSASRLQEDQ
 ERIITL

BASE COUNT 530 a 378 c 490 g 445 t 2 others
 ORIGIN

Query Match 30.1%; Score 576; DB 10; Length 1845;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1426; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 84 gtgtgagcgtatgaaatcagagagtgatgacgtcagactcgaagaagcagcgtacgcac 143
 |||||||
 Db 75 GTTGGCGGCGATGAGATGAGATCGAGAGTGCATCTCGAAGAAAGCAGCGCATCGCAC 134
 |||||||
 QY 144 ccaacccacatcaaggagactcgcgtcagcagcaatggaatgagatgctgtgtggcgc 203
 |||||||
 Db 135 CCACACCCACATCAAGGAGCTCGGCTTCGACGCCAATGGATGGCGATTGGCTGGCGGC 194
 |||||||
 QY 204 ggggttcgtggccagcagc 263
 |||||||
 Db 195 GGGGTTCTGGGCGCAGTGGCGCGCGCGCGAGGCGCGGCTGGCGGTCCACATGATTCC 254
 |||||||
 QY 264 ccaagaagaatgagc 323
 |||||||
 Db 225 CCAGAAAAAGATGGCGCGCGCGCGCGCTCTGCTGCTGGCTCGCGCGCGCGCGCAAGAC 314
 |||||||
 QY 324 ggcgtcagcgtcgc 383
 |||||||
 Db 315 GCGGCTTACCGCTCGGCAATAGCCAGAGCTCGGAGCAAGGCTCCCTTCTGCTTATGGT 374
 |||||||
 QY 384 aggatcagaagtgtactcctcgcaggtcagaagaactgaagtgctgtatggaataatccg 443
 |||||||
 Db 375 AGGATCAAGAAGTACTCTCTCGAGGTCAAGAAAACGTAGTGTCTGATGAAAAATTTCCG 434
 |||||||
 QY 444 tagagctaatggttgcgtataaagaagaacaaagaagttatgaagaagaagttactga 503
 |||||||
 Db 435 TAGACCTATGAGTGTGCTGCTATGAAGAAACAAAGAGGTTATGAAGAGGTTACTGA 494
 |||||||
 QY 504 acttccccagaagaaggtctgagatgatacaactggtgtgatatgcaaaagactagaact 563
 |||||||
 Db 495 ACTTTCGCCAGAGAGGCTGAGAGTACACACTGTGTGATATGCCAAGAAAGCTTACCTGT 554
 |||||||
 QY 564 aatcatcagcttaagactgtttaaaggagactaaagcaactgaagttagattcctaatta 623
 |||||||
 Db 555 AATCATGAGCTTAAAGACTGTTAAAGGAGCTAAGCACTGAAGTTAGATTCTTCAATTGA 614
 |||||||
 QY 624 tgaatccttgatcaagaagaagtgagcagtggtgtgattatatacatcgaaagctag 683
 |||||||
 Db 615 TGATGCTCTGATCAAGGAAAGGTGCGAGTGGGTGATGTTTATACATTGAAGCAAAATAG 674
 |||||||
 QY 684 tggagcagtgaaaagagttgtgtagatgtattcttcttcagagaatcagatcttgaagc 743
 |||||||
 Db 675 TGGACCAAGTGAAGAAAGTGGTAGATGTGATTTCTTTTGCTACAAATACATCTTGAAGC 734
 |||||||
 QY 744 tgaagagatgtttctatccccaaagtgtaagttccataagaagaaagaaatgtgcaaga 803
 |||||||
 Db 735 TGAAGAAATATGTTCTTATCCCAAGGTGAAGTCCATTAAGAAAAAATAATAGTGCAGGA 794
 |||||||
 QY 804 tgtcaacttcatgacacttgatgcaagcaaatgctcagcgcgaagaagtggtccaagtattc 863
 |||||||
 Db 795 TGTCACTTTCATGACCTTGTATGACGCAAAATGCTACGCCAAGAGTGGCCAAAGTATTTT 854
 |||||||
 QY 864 gtcccttatggccagatgagtgaaacccagaaagactgaaatccacgaaaaactcagcca 923
 |||||||
 Db 855 GTCCCTTATGGGCCAGATGATGAACCAACGAAAGATGAATATCCGAAAAAATACAGCCA 914
 |||||||
 QY 924 agaaatataaagtgtgataatagatalatcgaatgaagaagtgtcagaagcttgaactgg 983
 |||||||
 Db 915 AGAATTTAATTAAGGTGTAATATAGATATATGATGAAGAAATTCAGAGCTTGTACCTGG 974
 |||||||
 QY 984 tgtttgttcaattgtagtggtccacatgttgatatacgaatgttttctatctaaacg 1043
 |||||||

DEFINITION Oryza sativa chromosome 1 clone P0506B12, *** SEQUENCING IN
 PROGRESS ***, in ordered pieces.
 ACCESSION AP003271
 VERSION AP003271.1 GI:13027301
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0506B12.
 ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.
 REFERENCE 1 (sites)
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
 TITLE Oryza sativa nipponbare(GA5) genomic DNA, chromosome 1, PAC
 JOURNAL clome:P0506B12
 REFERENCE Published Only in Database (2001) In press
 AUTHORS 2 (bases 1 to 171809)
 TITLE Sasaki,T., Matsumoto,T. and Yamamoto,K.
 JOURNAL Direct Submission
 COMMENT Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
 Agricultural Resources, Rice Genome Research Program; Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan.
 (E-mail:tsasaki@affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1..171809
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0506B12"
 BASE COUNT 49272 a 36863 c 37265 g 48309 t 100 others
 ORIGIN
 Query Match 2.1%; Score 40; DB 83; Length 171809;
 Best Local Similarity 100.0%; Pred. No. 1.8e-11;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 688 gcagtgaaagagtgtagatgattcttctgctacagaatacagatctgaagctga 746
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 7696 GCAGTGAAAGAGTGTGATGATCTTTTGTGCTACAGATACGATCTGACGCTGA 7638
 RESULT 7
 LOCUS CNS06LVV 1173 bp DNA STS 10-JAN-2001
 DEFINITION T7 end of clone XAT0AA001H09 of library XAT0AA from strain CBS 4311
 of Saccharomyces servazzii, sequence tagged site.
 ACCESSION AL404705
 VERSION AL404705.1 GI:12166507
 KEYWORDS STS.
 SOURCE Saccharomyces servazzii.
 ORGANISM Saccharomyces servazzii.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 1 (bases 1 to 1173)
 Casaregola,S., Lepingle,A., Bon,E., Neuvéglise,C., Nguyen,H.,
 Artiguenave,F., Winkler,P. and Galliard,J.C.
 TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 7.
 JOURNAL Saccharomyces servazzii
 PUBMED FEBS Lett. 487 (1), 47-51 (2000)
 REFERENCE 1152882
 2 (bases 1 to 1173)

AUTHORS Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nicolas,C., Weslowski-Louvel,M.,
 Winkler,P. and Weissensbach,J.
 TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
 yeast species for molecular evolution studies(1)
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
 PUBMED 1152876
 REFERENCE 3 (bases 1 to 1173)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT This STS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.
 FEATURES
 source
 1..1173
 /organism="Saccharomyces servazzii"
 /strain="CBS 4311"
 /db_xref="taxon:27293"
 /clone="XAT0AA001H09"
 /clone_1bp="XAT0AA"
 /note="end : T7"
 STS 1..1173
 /note="part of mitochondrial DNA"
 /evidence="not_experimental"
 BASE COUNT 405 a 173 c 64 g 398 t 133 others
 ORIGIN
 Query Match 1.8%; Score 35; DB 53; Length 1173;
 Best Local Similarity 100.0%; Pred. No. 9.4e-09;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1878 wtaaaaaaataaaaaaataaaaaaataaaaaa 1912
 ||||||||||||||||||||||||||||||||||||||||
 Db 1104 WTAATAAAAAAAAAAAAAAAAAAAAAAAAAA 1138
 RESULT 8
 LOCUS I29928 55 bp DNA PAT 06-FEB-1997
 DEFINITION Sequence 41 from patent US 5578468.
 ACCESSION I29928
 VERSION I29928.1 GI:1820719
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE Unclassified.
 1 (bases 1 to 55)
 AUTHORS Pickup,D.J., Patel,D. and Antczak,J.B.
 TITLE Site-specific RNA cleavage
 JOURNAL Patent: US 5578468-A 41 26-NOV-1996;
 FEATURES Location/Qualifiers
 source
 1..55
 /organism="unknown"
 BASE COUNT 47 a 2 c 1 g 5 t
 ORIGIN
 Query Match 1.8%; Score 34; DB 10; Length 55;
 Best Local Similarity 100.0%; Pred. No. 2.8e-08;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
Db 15 TAAAAAAAAAAAAAAAAAAAAAAAAA 48

RESULT 9
S76508 71 bp DNA PRI 07-MAY-1993
LOCUS DIS8 (B) [hypervariable minisatellite MS32] [human, individual
DEFINITION AS89, Genomic, 71 nt].
ACCESSION S76508
VERSION S76508.1 GI:243383
KEYWORDS human individual AS89.
SOURCE Homo sapiens
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Monckton,D.G. and Jeffreys,A.J.
TITLE Minisatellite 'isocollie' discrimination in pseudohomozygotes by
JOURNAL single molecule PCR and variant repeat mapping
MEDLINE Genomics 11 (2), 465-467 (1991)
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 76508] from the original journal article.
This sequence comes from Figure 2.

FEATURES
source 1..71
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..71
/gene="DIS8"
/allele="B"

gene
BASE COUNT 66 a 0 c 0 g 5 t
ORIGIN

Query Match 1.8%; Score 34; DB 97; Length 71;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
Db 2 TAAAAAAAAAAAAAAAAAAAAAAAAA 35

RESULT 10
S76510 72 bp DNA PRI 07-MAY-1993
LOCUS DIS8 (B) [hypervariable minisatellite MS32] [human, individual
DEFINITION MACH, Genomic, 72 nt].
ACCESSION S76510
VERSION S76510.1 GI:243381
KEYWORDS human individual MACH.
SOURCE Homo sapiens
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Monckton,D.G. and Jeffreys,A.J.
TITLE Minisatellite 'isocollie' discrimination in pseudohomozygotes by
JOURNAL single molecule PCR and variant repeat mapping
MEDLINE Genomics 11 (2), 465-467 (1991)
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 76510] from the original journal article.
This sequence comes from Figure 2.

FEATURES
source 1..72
/organism="Homo sapiens"
/db_xref="taxon:9606"

gene
BASE COUNT 62 a 0 c 0 g 10 t
ORIGIN

Query Match 1.8%; Score 34; DB 97; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
Db 1 TAAAAAAAAAAAAAAAAAAAAAAAAA 34

RESULT 11
A08915 73 bp DNA PAT 02-SEP-1993
LOCUS H.sapiens (haplotype 3, allele MS32, isolate French, serial number
DEFINITION 8) minisatellite sequence.
ACCESSION A08915
VERSION A08915.1 GI:411837
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Jeffreys,A.J.
TITLE Extended nucleotide sequences
JOURNAL Patent: EP 0370719-A 96 30-MAY-1990;
IMPERIAL CHEMICAL INDUSTRIES PLC
location/Qualifiers

FEATURES
source 1..73
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 69 a 0 c 0 g 4 t
ORIGIN

Query Match 1.8%; Score 34; DB 97; Length 73;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
Db 1 TAAAAAAAAAAAAAAAAAAAAAAAAA 34

RESULT 12
A08908 77 bp DNA PAT 02-SEP-1993
LOCUS H.sapiens (haplotype 3, allele MS32, isolate Mormon, serial number
DEFINITION 10) minisatellite sequence.
ACCESSION A08908
VERSION A08908.1 GI:411830
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Jeffreys,A.J.
TITLE Extended nucleotide sequences
JOURNAL Patent: EP 0370719-A 91 30-MAY-1990;
IMPERIAL CHEMICAL INDUSTRIES PLC
location/Qualifiers

FEATURES
source 1..77
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 72 a 0 c 0 g 5 t
ORIGIN

Query Match 1.8%; Score 34; DB 9; Length 77;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taataaaataaaataaaataaaataaa 1912
|||||
Db 5 TAAATAAAAAAAAAAAAAAAAAAAAAA 38

RESULT 13
AX099452 79 bp DNA PAT 02-APR-2001
LOCUS Sequence 92 from Patent WO0119988.
ACCESSION AX099452
VERSION AX099452.1 GI:13538555
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 79)
AUTHORS Jacobs,K., McCooy,J.M., Lavallie,E.R., Collins-Racie,L.A., Evans,C.,
Metberg,D., Treacy,M., Bowman,M.R., Spaulding,V. and Agostino,M.J.
TITLE Secreted proteins and polynucleotides encoding them
JOURNAL Patent: WO 0119988-A 92 22-MAR-2001;
Genetics Institute, Inc. (US)
FEATURES Location/Qualifiers
source 1..79
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 64 a 2 c 3 g 10 t
ORIGIN

Query Match 1.8%; Score 34; DB 10; Length 79;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taataaaataaaataaaataaaataaa 1912
|||||
Db 27 TAAATAAAAAAAAAAAAAAAAAAAAAA 60

RESULT 14
A08907 80 bp DNA PAT 02-SEP-1993
LOCUS H.sapiens (haplotype 3, allele MS32, isolate English, serial number
DEFINITION 11) minisatellite sequence.
ACCESSION A08907
VERSION A08907.1 GI:411829
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 80)
AUTHORS Jeffreys,A.J.
TITLE Extended nucleotide sequences
JOURNAL Patent: EP 0370719-A 90 30-MAY-1990;
IMPERIAL CHEMICAL INDUSTRIES PLC
FEATURES Location/Qualifiers
source 1..80
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 72 a 0 c 0 g 8 t
ORIGIN

Query Match 1.8%; Score 34; DB 9; Length 80;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taataaaataaaataaaataaaataaa 1912
|||||
Db 7 TAAATAAAAAAAAAAAAAAAAAAAAAA 40

RESULT 15
AX098165/c 83 bp DNA PAT 30-MAR-2001
LOCUS Sequence 77 from Patent WO0118542.
DEFINITION AX098165
ACCESSION AX098165
VERSION AX098165.1 GI:13515205
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 83)
AUTHORS Lee,J., Thompson,P. and Lillie,J.
TITLE Identification, assessment, prevention, and therapy of ovarian
JOURNAL Patent: WO 0118542-A 77 15-MAR-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES Location/Qualifiers
source 1..83
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 12 a 10 c 13 g 47 t 1 others
ORIGIN

Query Match 1.8%; Score 34; DB 10; Length 83;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taataaaataaaataaaataaaataaa 1912
|||||
Db 64 TAAATAAAAAAAAAAAAAAAAAAAAAA 31

Search completed: November 13, 2001, 09:53:05
Job time: 6926 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2001, 08:44:44 ; Search time 144.47 Seconds
(without alignments)
8310.015 Million cell updates/sec

Title: US-09-589-510-3

Perfect score: 1912
Sequence: 1 acccaagcttcgcaattt.....aaaaaaaaaaaaaaaaaa 1912

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N.Geneseq_0601.*

1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1912	100.0	1912	22	MAIZE RuVB ortholog
2	1080	56.5	1869	22	MAIZE RuVB ortholog
3	576	30.1	1845	22	MAIZE RuVB ortholog
4	94	4.9	1886	22	MAIZE RuVB ortholog
5	94	4.9	1898	22	MAIZE RuVB ortholog
6	94	4.9	1898	22	MAIZE RuVB ortholog
7	34	1.8	69	18	Human secreted protein
8	34	1.8	79	19	Human secreted protein
9	34	1.8	79	19	Human secreted protein
10	34	1.8	80	19	Human secreted protein
11	34	1.8	81	21	Human secreted protein

ALIGNMENTS

12	34	1.8	85	19	AAV05720	Nucleotide sequence
13	34	1.8	86	12	AAO11760	Self-complementary
14	34	1.8	89	21	AACT8380	Human cancer assoc
15	34	1.8	90	12	AAO11762	Self-complementary
16	34	1.8	90	21	AAE18395	Lung cancer associ
17	34	1.8	95	21	AAE16473	Human prostate can
18	34	1.8	98	20	AAAX00174	Porcine reproductive
19	34	1.8	99	18	AAE91300	Human M97-2 secret
20	34	1.8	101	18	AAV00420	3' fragment of clo
21	34	1.8	111	21	AAE98625	Human colon cancer
22	34	1.8	114	17	AAE39472	Growth regulatory
23	34	1.8	114	19	AAV37490	Human growth regul
24	34	1.8	120	14	AAO52429	Human subtelomeric
25	34	1.8	120	16	AAO87011	Subtelomeric cDNA
26	34	1.8	120	19	AAE96338	Staphylococcus aur
27	34	1.8	140	18	AAE76781	Immunodominant fra
28	34	1.8	141	16	AAE06012	Homo sapiens clone
29	34	1.8	144	19	AAV21239	Human secreted pro
30	34	1.8	144	20	AAE77331	Capture probe for
31	34	1.8	191	17	AAE113475	Nucleotide sequenc
32	34	1.8	198	21	AAE64592	3' fragment of clo
33	34	1.8	201	18	AAV00418	Human colon cancer
34	34	1.8	202	21	AAE98563	Human gene express
35	34	1.8	208	20	AAE14464	Murine secreted pr
36	34	1.8	208	20	AAE77547	3' nucleotide sequ
37	34	1.8	216	19	AAV09114	Murine secreted pr
38	34	1.8	216	20	AAE77545	Sequence secreted pr
39	34	1.8	216	21	AAE59487	Downstream sequenc
40	34	1.8	222	14	AAO46071	Human lung tumour
41	34	1.8	222	22	AAE86777	Human lung tumour
42	34	1.8	229	22	AAE68277	Human colon cancer
43	34	1.8	231	21	AAE98505	Human colon cancer
44	34	1.8	234	21	AAE98428	Human prostate can
45	34	1.8	236	21	AAE16372	

RESULT 1

AAO2567	standard; cDNA; 1912 bp.
ID	AAO2567
AC	AAO2567
XX	02-MAY-2001 (first entry)
DT	
XX	MAIZE RuVB orthologue #2 cDNA.
DE	
XX	MAIZE; RuVB orthologue; branch migration; heteroduplex extension;
XX	KW homologous recombination; transformation; transgenic plant; ss.
OS	Zea mays.
XX	
FT	key
FT	location/Qualifiers
FT	94..1461
FT	/*tag= a
XX	/product= "Maize RuVB orthologue protein #2"
PN	WO200105975-A1.
XX	
PD	25-JAN-2001.
XX	
PF	13-JUN-2000; 2000WO-US16271.
XX	
PR	16-JUL-1999; 99US-0144112.
XX	
PA	(PTON-) PIONEER HT-BRED INT INC.
XX	
PI	Mahajan PB;
XX	
DR	WPI: 2001-159537/16.
DR	P-PSDB; AAY72563.

[illegible]

Db	601	ctggaagcttagatcttccaaatttaagatgctcttgatcaaggaaaagctggcagtgtagat	660
Qy	661	gttatcatcatcgaaagcaaatagtcgagccgtgaaagaatctgttagatgcatctctt	720
Db	661	gttatcatcatcgaaagcaaatagtcgagccgtgaaagaatctgttagatgcatctctt	720
Qy	721	gctacagaatacgaatcttgaaagctgaaagatagttcttaccaccaagctgaagtcac	780
Db	721	gctacagaatacgaatcttgaaagctgaaagatagttcttaccaccaagctgaagtcac	780
Qy	781	aagaaaaaagaatttggcaggaatgatacacttcacagacttgatgacgaatgctcag	840
Db	781	aagaaaaaagaatttggcaggaatgatacacttcacagacttgatgacgaatgctcag	840
Qy	841	ccacaagctggccaagatatttgccttatggccagaatgatacgaacacgaagaagct	900
Db	841	ccacaagctggccaagatatttgccttatggccagaatgatacgaacacgaagaagct	900
Qy	901	gaatcacccgaaaaaactacgcaccaagaataataaagctggttaaatagatatactga	960
Db	901	gaatcacccgaaaaaactacgcaccaagaataataaagctggttaaatagatatactga	960
Qy	961	ggaattgcaagccttgcacctgtgttttcttcaatgataagagttccaaatgctgatatc	1020
Db	961	ggaattgcaagccttgcacctgtgttttcttcaatgataagagttccaaatgctgatatc	1020
Qy	1021	gaattttttcttacttaaacccgcacatggagagcccaattacaccaatgtgttact	1080
Db	1021	gaattttttcttacttaaacccgcacatggagagcccaattacaccaatgtgttact	1080
Qy	1081	gctacaaataggggataatgataatgataaggaactgatactgaacagttccaaatgata	1140
Db	1081	gctacaaataggggataatgataatgataaggaactgatactgaacagttccaaatgata	1140
Qy	1141	ccggtgggactcttctaagataagctggtgattatctcgaaacgaacataatggcctactag	1200
Db	1141	ccggtgggactcttctaagataagctggtgattatctcgaaacgaacataatggcctactag	1200
Qy	1201	atgatacgaatattggtataccgagcaacgtggaggaattgatatgatagaataagaagt	1260
Db	1201	atgatacgaatattggtataccgagcaacgtggaggaattgatatgatagaataagaagt	1260
Qy	1261	ctgctatttaaggcgaatccgagacgaacactctttagacatgcatataatgata	1320
Db	1261	ctgctatttaaggcgaatccgagacgaacactctttagacatgcatataatgata	1320
Qy	1321	tcacctgcccagcgtgtgtcttaagaactaattggaagagagaaaaatctgcaagctgatactc	1380
Db	1321	tcacctgcccagcgtgtgtcttaagaactaattggaagagagaaaaatctgcaagctgatactc	1380
Qy	1381	gaggaaatcagttgggctctatttggatgccaatctcctggcctcgagctcccaagagaa	1440
Db	1381	gaggaaatcagttgggctctatttggatgccaatctcctggcctcgagctcccaagagaa	1440
Qy	1441	caagaaagatacatcacactagatttggatctcctgtcgtggaagctcgaagagaaatgta	1500
Db	1441	caagaaagatacatcacactagatttggatctcctgtcgtggaagctcgaagagaaatgta	1500
Qy	1501	gttgccagctcggaaagtcatctagctgcatctgatacgtctcaagttcatagtctactgg	1560
Db	1501	gttgccagctcggaaagtcatctagctgcatctgatacgtctcaagttcatagtctactgg	1560
Qy	1561	tccttgagcgaagacaatttcgggggggaagggcttgaatttgcagctgcgtctgtgta	1620
Db	1561	tccttgagcgaagacaatttcgggggggaagggcttgaatttgcagctgcgtctgtgta	1620
Qy	1621	gttcccaaraaagaactgtgtctccgagatactgtcgttcaacgaactgtctcgattaga	1680
Db	1621	gttcccaaraaagaactgtgtctccgagatactgtcgttcaacgaactgtctcgattaga	1680
Qy	1681	ttgttgcacggctgcaggaattggcgtgtgttcttctatctgtcatcgtgtccggagt	1740
Db	1681	ttgttgcacggctgcaggaattggcgtgtgttcttctatctgtcatcgtgtccggagt	1740

Db 1681 ttggtaccggtgacgaattccgtgtgtttttattcttcacatcggtccggagt 1740
QY 1741 ctgctccacaggttgatgtgcccgaacctactcttgtaaccatgataatgataag 1800
Db 1741 ctgctccacaggttgatgtgcccgaacctactcttgtaaccatgataatgataag 1800
QY 1801 attcttaacagaatgcgaactgtgcttattattcttaaatgltccataaagcataag 1860
Db 1801 attcttaacagaatgcgaactgtgcttattattcttaaatgltccataaagcataag 1860
QY 1861 aatgttttccacaacmtwtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912
Db 1861 aatgttttccacaacmtwtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912
RESULT 2
AAD02570
ID AAD02570 standard; cDNA; 1869 BP.
XX AAD02570:
AC AAD02570:
XX
DT 02-MAY-2001 (first entry)
XX
DE Maize RuVB orthologue #5 cDNA.
XX
KM Maize: RuVB orthologue; branch migration; heteroduplex extension;
KW homologous recombination; transformation; transgenic plant; ss.
XX
OS Zea mays.
XX
FT Key Location/Qualifiers
FT CDS 64..1380
FT /tag= a
FT /product= "Maize RuVB orthologue protein #5"
XX
PN W0200105975-A1.
XX
PD 25-JAN-2001.
XX
PF 13-JUN-2000; 2000MO-US16271.
XX
PR 16-JUL-1999; 99US-0144112.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
XX Mahajan PB;
XX
DR WPI: 2001-159537/16.
DR P-PSDB: AAY72566.
XX
PT Novel maize RuVB nucleic acid useful for modulating levels of maize
PT RuVB in plants, as probes or amplification primers in the detection,
PT quantitation or isolation of gene transcripts -
XX
PS Claim 1; Page 80-82; 87pp; English.
XX
CC The present sequence is a Zea mays RuVB orthologue #5 cDNA. RuVB along
CC with RuVB catalyses the branch migration process, also known as
CC heteroduplex extension, in homologous recombination. RuVB is used for
CC the control of homologous recombination or transformation efficiency in
CC transgenic plants. The RuVB nucleotide may be used as probes or
CC amplification primers for detecting, quantifying or isolating gene
CC transcripts, in detecting deficiencies in the mRNA level during screening
CC for desired transgenic plants, for detecting gene mutations, for
CC monitoring upregulation of expression or changes in enzyme activity, for
CC detecting any number of allelic variants, orthologues or paralogues of
CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
CC also be used for recombinant expression of its encoded polypeptide, or
CC for use as immunogen in preparing and/or screening of antibodies, and in
CC sense or antisense suppression of one or more genes in a host cell,
CC tissue or plant. The RuVB proteins may be used in assays to agonise or
CC antagonise the enzyme function, or as immunogens or antigens for
CC screening antibodies.

XX
SQ Sequence 1869 BP; 539 A; 373 C; 486 G; 471 T; 0 other;
Query Match 56.5%; Score 1080; DB 22; Length 1869;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1130; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 337 ggcataagcccaagagatctgcgcagcaagaagttcccttcgtctcctatgtagagatgaagatg 396
Db 256 ggcataagcccaagagatctgcgcagcaagaagttcccttcgtctcctatgtagagatgaagatg 315
QY 397 tactctcgagaggtcaagaactaagtgctgataaggaataatlcgtatagatagatg 456
Db 316 tactctcgagaggtcaagaactaagtgctgataaggaataatlcgtatagatagatg 375
QY 457 ttgcgtataaaggaataaagaaggttataaggaagagagttactgaacttccccaagaa 516
Db 376 ttgcgtataaaggaataaagaaggttataaggaagagagttactgaacttccccaagaa 435
QY 517 gaggtctgagagatgaactggtggtatataagcaaaagcatagcatgataatcatcagctta 576
Db 436 gaggtctgagagatgaactggtggtatataagcaaaagcatagcatgataatcatcagctta 495
QY 577 aagactgttaaaaggagactgaactgaactgaagttagatttcttaattatgatagtctgatac 636
Db 496 aagactgttaaaaggagactgaactgaactgaagttagatttcttaattatgatagtctgatac 555
QY 637 aagaaaaggtggtcaggtggtgataatatacatcgaagcaaatagtgtgagcagtgataa 696
Db 556 aagaaaaggtggtcaggtggtgataatatacatcgaagcaaatagtgtgagcagtgataa 615
QY 697 aaggttgtagatgtagatcttcttcctacagaatacagctcttgaaagcgtgaagatgatt 756
Db 616 aaggttgtagatgtagatcttcttcctacagaatacagctcttgaaagcgtgaagatgatt 675
QY 757 cctatccccaaggtgaaagtcataagaataaagaatgtgtcagaagtgtcaacttcac 816
Db 676 cctatccccaaggtgaaagtcataagaataaagaatgtgtcagaagtgtcaacttcac 735
QY 817 gaacttgatgcagcaaatgtctcagcccaaggtgtgccaagatatttgccttatgagc 876
Db 736 gaacttgatgcagcaaatgtctcagcccaaggtgtgccaagatatttgccttatgagc 795
QY 877 cagatgataaaccacgaagaactgaataatcacgcgaataaactacgcgaagaataataaag 936
Db 796 cagatgataaaccacgaagaactgaataatcacgcgaataaactacgcgaagaataataaag 855
QY 937 gtggtataatagatatacgaatgaagaatgtcagaagctgtgtaacctgtgtttgttcatt 996
Db 856 gtggtataatagatatacgaatgaagaatgtcagaagctgtgtaacctgtgtttgttcatt 915
QY 997 gatgaagttccacatgttgatatacgaatgttttcttacttaacacgtgcattggaagc 1056
Db 916 gatgaagttccacatgttgatatacgaatgttttcttacttaacacgtgcattggaagc 975
QY 1057 ccaattacccaatcgtgtgacttgctacaataggggataatgttaatgtaaggaagact 1116
Db 976 ccaattacccaatcgtgtgacttgctacaataggggataatgttaatgtaaggaagact 1035
QY 1117 gatatacaaatgcacatgatacgaatcggtgtgatactctctagataagctgtgtattcgg 1176
Db 1036 gatatacaaatgcacatgatacgaatcggtgtgatactctctagataagctgtgtattcgg 1095
QY 1177 acgagaacatagggcctctagatagatagacagatatgtgctatccgagcaacaagtggag 1236
Db 1096 acgagaacatagggcctctagatagatagacagatatgtgctatccgagcaacaagtggag 1155
QY 1237 gaatgatataatgaataaagaagcttctgttattttagggagatcggaaagcagacatct 1296
Db 1156 gaatgatataatgaataaagaagcttctgttattttagggagatcggaaagcagacatct 1215
QY 1297 ttgagaacatgctatcgaatgataatcactgcagcgtgtgtctcaaaagcaatgtgaaga 1356

```

Db      1216  ttgagatgcttcttcaatgatataccgacgctggtccaaagactaatggaaga 1275
QY      1357  gagaataatctgcaaggctgactcgcgaggaagtcagtggtgctctatttgatgcgaatcc 1416
Db      1276  gagaataatctgcaaggctgactcgcgaggaagtcagtggtgctctatttgatgcgaatcc 1335
QY      1417  tcggtcgcgtgctcgcgaggaacaaagatacatcactagattgg 1467
Db      1336  tcggtcgcgtgctcgcgaggaacaaagatacatcactagattgg 1386

RESULT  3
AAD02566
ID      AAD02566 standard; cDNA; 1845 BP.
XX
AC      AAD02566;
XX
DE      02-MAY-2001 (first entry)
XX
DE      Maize Ruvb orthologue #1 cDNA.
XX
KW      Maize; Ruvb orthologue; branch migration; heteroduplex extension;
XX      homologous recombination; transformation; transgenic plant; ss.
XX      Zea mays.
XX
FH      Key      Location/Qualifiers
FT      CDS      85..1452
FT      /tag=a
FT      /product= "Maize Ruvb orthologue protein #1"
XX
PN      W020105975-A1.
XX
PD      25-JAN-2001.
XX
PE      13-JUN-2000; 2000WO-US16271.
XX
PR      16-JUL-1999; 99US-0144112.
XX
PA      (PION-) PIONEER HI-BRED INT INC.
XX
PI      Mahajan PB;
XX
DR      MPI: 2001-159537/16.
DR      P-PSDB: AAY72562.
XX
PT      Novel maize Ruvb nucleic acid useful for modulating levels of maize
PT      Ruvb in plants, as probes or amplification primers in the detection,
PT      quantitation or isolation of gene transcripts -
XX
PS      Claim 1; Page 67-69; 87pp; English.
XX
CC      The present sequence is a Zea mays Ruvb orthologue #1 cDNA. Ruvb along
CC      with Ruva catalyses the branch migration process, also known as
CC      heteroduplex extension, in homologous recombination. Ruvb is used for
CC      the control of homologous recombination or transformation efficiency in
CC      transgenic plants. The Ruvb nucleotide may be used as probes or
CC      amplification primers for detecting, quantifying or isolating gene
CC      transcripts, in detecting deficiencies in the mRNA level during gene
CC      transfer for desired transgenic plants, for detecting gene mutations, for
CC      monitoring upregulation of expression or changes in enzyme activity, for
CC      detecting any number of allelic variants, orthologues or paralogues of
CC      the gene, or for site directed mutagenesis in eukaryotic cells. It may
CC      also be used for recombinant expression of its encoded polypeptide, or
CC      for use as immunogen in preparing and/or screening of antibodies, and in
CC      sense or antisense suppression of one or more genes in a host cell,
CC      tissue or plant. The Ruvb proteins may be used in assays to agonise or
CC      antagonise the enzyme function, or as immunogens or antigens for
CC      screening antibodies.
XX
SQ      Sequence 1845 BP; 530 A; 378 C; 490 G; 445 T; 2 other:

```

```

Query Match      30.1%; Score 576; DB 22; Length 1845;
Best Local Similarity 98.8%; Pred. No. 2.2e-214;
Matches 1426; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      84  gttggcgcgataagatcagagagtgcaatgcagctcgaagaagcagcgcac 143
Db      75  gttggcgcgataagatcagagagtgcaatgcagctcgaagaagcagcgcac 134
QY      144  ccaaccccaatcaaggactcgcctcgcagccatggatggatggatggatgg 203
Db      135  ccaaccccaatcaaggactcgcctcgcagccatggatggatggatggatgg 194
QY      204  ggggttcgtggcgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 203
Db      195  ggggttcgtggcgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 254
QY      264  ccagaagaagaatggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 323
Db      255  ccagaagaagaatggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 314
QY      324  ggcgtacgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 383
Db      315  ggcgtacgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 374
QY      384  agatcagaagtgctacccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 443
Db      375  agatcagaagtgctacccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 434
QY      444  taagctatagtttgctataaagaaacaaagagttatgaagagagttactga 503
Db      435  taagctatagtttgctataaagaaacaaagagttatgaagagagttactga 494
QY      504  acttcccaagagagctgaagctacactgtgtatatagcacaaagcattagc 563
Db      495  acttcccaagagagctgaagctacactgtgtatatagcacaaagcattagc 554
QY      564  aatcatagcttaaaagctgttaaagggaacgaacgaagttgattctcaatta 623
Db      555  aatcatagcttaaaagctgttaaagggaacgaacgaagttgattctcaatta 614
QY      624  tgatgctcgtatcaagaagaagtggtgcaatgtatatacatcgaagaacatag 683
Db      615  tgatgctcgtatcaagaagaagtggtgcaatgtatatacatcgaagaacatag 674
QY      684  tggagcagtgaaagagtggtgcaatgtatatacatcgaagaacatag 743
Db      675  tggagcagtgaaagagtggtgcaatgtatatacatcgaagaacatag 734
QY      744  tgaagatgttctatcccaagtggaatgcataaagaagaagaatgtgcaaga 803
Db      735  tgaagatgttctatcccaagtggaatgcataaagaagaagaatgtgcaaga 794
QY      804  tgtacacttcataagccttgatgcgaatgtcagccacaaagtggtgcaagaatt 863
Db      795  tgtacacttcataagccttgatgcgaatgtcagccacaaagtggtgcaagaatt 854
QY      864  gtccttatggcgcgaatgtatgaacccaggaagaagctgaatccgcgaagaac 923
Db      855  gtccttatggcgcgaatgtatgaacccaggaagaagctgaatccgcgaagaac 914
QY      924  aagaatataagtggtgaatatacatcgaatgaagaagttcagaagcttgatc 983
Db      915  aagaatataagtggtgaatatacatcgaatgaagaagttcagaagcttgatc 974
QY      984  tgttttcatttgatgaagtcacatgttgatatacgaatgtttctatacctaac 1043
Db      975  tgttttcatttgatgaagtcacatgttgatatacgaatgtttctatacctaac 1034
QY      1044  tgcattgagagcccatatataccaatcgtatatactgctacaaatagggaatata 1103
Db      1035  tgcattgagagcccatatataccaatcgtatatactgctacaaatagggaatata 1094

```


XX PS Disclosure: Page 77-79; 87pp; English.

CC The present sequence is a *Zea mays* RuVb orthologue #4 cDNA. RuVb along
 CC with RuVb catalyses the branch migration process, also known as
 CC heteroduplex extension, in homologous recombination. RuVb is used for
 CC the control of homologous recombination or transformation efficiency in
 CC transgenic plants. The RuVb nucleotide may be used as probes or
 CC amplification primers for detecting, quantifying or isolating gene
 CC transcripts, in detecting deficiencies in the mRNA level during screening
 CC for desired transgenic plants, for detecting gene mutations, for
 CC monitoring upregulation of expression or changes in enzyme activity, for
 CC detecting any number of allelic variants, orthologues or paralogues of
 CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
 CC also be used for recombinant expression of its encoded polypeptide, or
 CC for use as immunogen in preparing and/or screening of antibodies, and in
 CC sense or antisense suppression of one or more genes in a host cell,
 CC tissue or plant. The RuVb proteins may be used in assays to agonise or
 CC antagonise the enzyme function, or as immunogens or antigens for
 CC screening antibodies.

XX SQ Sequence 1898 BP; 532 A; 413 C; 521 G; 432 T; 0 other;

Query Match 4.9%; Score 94; DB 22; Length 1898;
 Best Local Similarity 100.0%; Pred. No. 1.4e-27;
 Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 gaggtcaagaacacgaggtgctgatagtgaaattccgtagagctataggttgcgtata 465
 DB 481 gaggtcaagaacacgaggtgctgatagtgaaattccgtagagctataggttgcgtata 540

QY 466 aaggaacacacagaggttcatgaaggaagaggtta 499
 DB 541 aaggaacacacagaggttcatgaaggaagaggtta 574

RESULT 6
 AAT88081
 ID AAT88081 standard; cDNA; 69 BP.

AC AAT88081;
 AT 14-MAY-1998 (first entry)
 DE 3' portion of cDNA clone encoding secreted protein AK296.
 KW Human; secreted protein; research; treatment; AM533;
 KW 3' portion; ds.
 OS Homo sapiens.
 OS WO9739122-A2.
 PN 23-OCT-1997.
 PD 11-APR-1997; 97WO-US06042.
 PF 12-APR-1996; 96US-0631184.
 PR (MURRO-) MURRO PHARM INC.
 PA Theoharides TC;
 PI WPI; 1997-526459/48.
 DR Human and murine secreted proteins - useful to research or treat
 PT diseases or disorders related to their function
 PS Disclosure: Page 88; 140pp; English.

XX CC The present sequence is the 3' portion of a cDNA clone encoding a
 CC human secreted protein, which may have nutritional uses, or

CC cytokine and cell proliferation/differentiation, immune stimulating
 CC or suppressing, haematopoiesis regulating, tissue growth,
 CC activin/inhibin, chemotactic/chemokinetic, haemostatic and
 CC thrombolytic, receptor/ligand, anti-inflammatory or tumour
 CC inhibition activities. It can also be used to research or treat
 CC diseases/disorders related to its function.

CC The partial cDNA clone API62 was 1st isolated from a human adult
 CC placenta cDNA library. The partial cDNA clones AK931, AM610, AM340,
 CC AM882, AK447, AK583, AK533 and AK296 were 1st isolated from a human
 CC foetal kidney cDNA library. The partial cDNA clones H617 and B89
 CC were 1st isolated from a human peripheral blood monocyte cell (Th1
 CC or Th2) cDNA library. The partial cDNA clone AM191 was 1st isolated
 CC from a human ovary (PA-1 teratocarcinoma) cDNA library. The partial
 CC cDNA clones AT211, AT205 and AT319 were 1st isolated from a human
 CC lymphocyte and dendritic cell cDNA library. The partial cDNA clones
 CC AS34 and AS32 were 1st isolated from a human foetal brain cDNA
 CC library. The partial cDNA clone AR260 was 1st isolated from a human
 CC adult retina cDNA library. The partial cDNA clones K640 and K39
 CC were 1st isolated from a murine bone marrow (stromal cell line
 CC FCM-4) cDNA library.

XX SQ Sequence 69 BP; 48 A; 6 C; 4 G; 9 T; 2 other;

Query Match 1.8%; Score 34; DB 18; Length 69;
 Best Local Similarity 100.0%; Pred. No. 0.00042;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 taataaaaaaaaaaaaaaaaaaaaaaaaaa 1912
 DB 5 taataaaaaaaaaaaaaaaaaaaaaaaaaa 38

RESULT 7
 AAV02148
 ID AAV02148 standard; cDNA; 69 BP.

AC AAV02148;
 AT 12-MAY-1998 (first entry)
 DE Human secreted protein AK296 3' portion including the polyA tail.
 KW Human; secreted protein; ATCC 98026; cytokine; immunomodulation;
 KW cell proliferation; differentiation; regulation; ds.
 OS Homo sapiens.
 OS WO9739123-A2.
 PN 23-OCT-1997.
 PD 14-APR-1997; 97WO-US06139.
 PF 18-APR-1996; 96US-0634325.
 PR (GENMY) GENETICS INST INC.
 PA Jacobs K, Lavallic ER, McCoy JM, Merberg D, Racie LA;
 PI Spaulding V;
 PI WPI; 1997-526460/48.
 DR New secreted proteins encoded clones present in ATCC 98026 -
 PT possibly having cytokine, cell proliferation/differentiation
 PT regulating, immunomodulating and many other activities
 PS Disclosure: Page 88; 139pp; English.

XX CC The present sequence encodes a portion of a novel human secreted protein
 CC deposited under accession number ATCC 98026. The secreted protein can be
 CC used to determine biological activity, to raise antibodies, as tissue
 CC markers, to isolate cognate ligands or receptors, to identify agents

CC that modulate their interactions and as nutritional supplements. It may
 CC also have a very wide range of biological activities although no
 CC evidence for any is provided in the specification. Typical of these are
 CC cytokine, cell proliferation/differentiation modulating activity or
 CC induction of other cytokines; immunostimulating/immunosuppressant
 CC activities (e.g. for treating human immunodeficiency virus infection,
 CC cancer, autoimmune diseases and allergy); regulation of haematopoiesis
 CC (e.g. for treating anaemia or as adjunct to chemotherapy); stimulation
 CC of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for
 CC treating wounds, periodontal disease, neurological diseases stroke,
 CC fibrosis); inhibition or stimulation of follicle stimulating hormone
 CC (for control of fertility); chemotactic and chemokinetic activities
 CC (e.g. for treating infections, tumours); haemostatic or thrombolytic
 CC activity (e.g. for treating haemophilia, cardiac infarction etc.);
 CC anti-inflammatory activity (e.g. for treating septic shock, Crohn's
 CC disease); as antimicrobials; for treating psoriasis or other
 CC hyperproliferative disease; for regulation of metabolism, behaviour, and
 CC many others. Also contemplated is the use of the corresponding nucleic
 CC acid in gene therapy procedures.

SQ Sequence 69 BP; 48 A; 6 C; 4 G; 9 T; 2 other;

Query Match 1.8%; Score 34; DB 18; Length 69;
 Best Local Similarity 100.0%; Pred. No. 0.00042;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taataaaaaaaaaaaaaaaaaaaaaa 1912
 |||
 Db 5 taataaaaaaaaaaaaaaaaaaaaaa 38

RESULT 8

AAV32414
 ID AAV32414 standard; cDNA: 79 BP.

XX AAV32414;

XX 13-OCT-1998 (first entry)

DE Homo sapiens clone BD441_2 3' region.

XX Secreted protein; clone; BD441_2; ds.

OS Homo sapiens.

XX WO9822501-A2.

XX 28-MAY-1998.

PF 19-NOV-1997; 97WO-US21123.

XX 17-NOV-1997; 97US-0971786.

PR 20-NOV-1996; 96US-0752912.

PR 14-FEB-1997; 97US-0800826.

XX (GEMV) GENETICS INST INC.

PI Aostino MJ, Jacobs K, Lavallie ER, McCoy JM, Werberg D;
 PI Racie LA, Spaulding V, Treacy M;
 WIPI; 1998-312414/27.

PT New nucleic acid encoding secreted protein from human cells -
 PT potentially useful, e.g. as immuno-modulators, antitumor agents,
 PT promoters of tissue growth, haemostatic and thrombolytic agents

XX Claim 13; Page 61; 93pp; English.

CC The sequence is that of the 3' region of clone BD441_2 which
 CC may be of use as a tissue/molecular weight marker; for chromosome
 CC identification, to identify possible genetic disorders, to isolate
 CC new related DNA or as a source of primers for PCR.

XX Sequence 79 BP; 64 A; 2 C; 3 G; 10 T; 0 other;

Query Match 1.8%; Score 34; DB 19; Length 79;
 Best Local Similarity 100.0%; Pred. No. 0.00041;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taataaaaaaaaaaaaaaaaaaaaaa 1912
 |||
 Db 27 taataaaaaaaaaaaaaaaaaaaaaa 60

RESULT 9

AAV26706
 ID AAV26706 standard; cDNA: 79 BP.

XX AAV26706;

XX 11-SEP-1998 (first entry)

DE Human novel secreted protein clone BF290_1i 3'-end DNA.

XX Secreted protein; human; marker; cytokine; immune stimulant; suppressor;

KW autoimmune disease; regulator; activin; inhibin; inhibitor; chemotactic;

KW chemokinetic; haemostatic; thrombocyte; tumour; anti-inflammatory; ds.

XX Homo sapiens.

OS WO9814470-A2.

XX 09-APR-1998.

PF 03-OCT-1997; 97WO-US18032.

PR 04-OCT-1996; 96US-0725885.

XX (GEMV) GENETICS INST INC.

PI Jacobs K, Lavallie ER, McCoy JM, Werberg D, Racie LA;
 PI Spaulding V, Treacy M;
 WIPI; 1998-240017/21.

PT Compositions containing isolated proteins encoded by specific
 PT nucleic acids - useful e.g. as anti-inflammatory, immune stimulant
 PT or suppressor, etc.

PS Disclosure; Page 94; 117pp; English.

CC This sequence encodes the 3'-end of a novel secreted protein from clone
 CC BF290_1i isolated from a human fetal brain cDNA library. Such secreted
 CC proteins can be used for, e.g. research purposes (as markers for
 CC tissues, molecular weight markers for gels, primers and probes), for
 CC nutrition (as C, N or carbohydrate source), as a cytokine for cell
 CC proliferation and differentiation activity, as immune stimulants or
 CC suppressors, e.g. for viral, bacterial or fungal infections, for
 CC autoimmune diseases such as multiple sclerosis or systemic lupus
 CC erythematosus, to regulate haematopoiesis, for tissue growth, as an
 CC activin or inhibin, or having chemotactic or chemo-kinetic, haemostatic
 CC and thrombocytic, receptor/ligand, anti-inflammatory or tumour inhibitory
 CC activities.

SQ Sequence 79 BP; 73 A; 0 C; 0 G; 6 T; 0 other;

Query Match 1.8%; Score 34; DB 19; Length 79;
 Best Local Similarity 100.0%; Pred. No. 0.00041;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taataaaaaaaaaaaaaaaaaaaaaa 1912
 |||
 Db 7 taataaaaaaaaaaaaaaaaaaaaaa 40

```

RESULT 10
AAV37197/c
ID AAV37197 standard; DNA; 80 BP.
XX
AC AAV37197;
XX
DT 14-SEP-1998 (first entry)
XX
DE Oligonucleotide sequence of the specification.
XX
KM Genotype; phenotype; molecular evolutionary engineering;
XX functional biopolymer; virus; ss.
OS Synthetic.
XX
PN WO9816636-A1.
XX
PD 23-APR-1998.
XX
PF 17-OCT-1997; 97WO-JP03766.
XX
PR 17-OCT-1996; 96JP-0274855.
XX
PA (MITU ) MITSUBISHI CHEM CORP.
XX
PI Fusimi Y, Miyamoto E, Nemoto N, Yanagawa H;
XX
DR WPI; 1998-261039/23.
XX
PT Virus containing nucleic acid and protein sections - for use in
XX modification and creation of functional bio-polymers such as
XX enzymes, antibodies and ribozyme(s)
XX
PS Example 1; Page 44; 68pp; Japanese.
XX
CC The present sequence is used in the course of the invention. The
XX specification describes a molecule for bringing together genotype with
XX phenotype (in vitro virus). The molecule contains a nucleic acid
XX fragment having a base sequence corresponding to a genotype, covalently
XX bonded to a protein fragment containing a protein participating in
XX phenotype expression, the 3'-end of the nucleic acid part being bonded
XX to the C-terminus of the protein part via a puromycin moiety. The nucleic
XX acid fragment preferably consists of RNA corresponding to the gene (free
XX from a termination codon), a spacer (such as two-stranded DNA), a peptide
XX adapter (such as a DNA-tRNA hybrid containing an anticodon corresponding
XX to the terminator codon of the gene) and a puromycin-containing cap
XX capable of binding to an amino acid residue. Translation of the virus is
XX performed in a non-cellular (preferably ribosomal) system (e.g. E. coli
XX ribosome), the protein synthesised by the translation attaching to the
XX puromycin cap, resulting in the complete in vitro virus structure. The
XX method is used in molecular evolutionary engineering to optimise function
XX of a functional biopolymer such as an enzyme, antibody or ribozyme, or
XX to generate new functionality. The virus can be used for the functional
XX optimisation of nucleic acid or protein sequences.
XX
SQ Sequence 80 BP; 7 A; 10 C; 1 G; 62 T; 0 other;

Query Match 1.8%; Score 34; DB 19; Length 80;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taataaaaaaaaaaaaaaaaaaaaaa 1912
    ||||||||||||||||||||||||||||
DB 64 TAAAAAAAAAAAAAAAAAAAAAAAAA 31

RESULT 11
AAC13249
ID AAC13249 standard; cDNA; 81 BP.
XX

```

```

AC AAC13249;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 17324.
XX
KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 17324; 71bp + CD-ROM; English.
XX
XX
CC The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
SQ Sequence 81 BP; 71 A; 2 C; 2 G; 6 T; 0 other;

Query Match 1.8%; Score 34; DB 21; Length 81;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taataaaaaaaaaaaaaaaaaaaaaa 1912
    ||||||||||||||||||||||||||||
DB 25 taataaaaaaaaaaaaaaaaaaaaaa 58

RESULT 12
AAV05720
ID AAV05720 standard; cDNA; 85 BP.
XX
AC AAV05720;
XX
DT 05-JUN-1998 (first entry)
XX
DE Nucleotide sequence of the 3' portion from clone AZ302.1.
XX
KM Colon; secreted protein; EST sequence; homology; antibody;
XX immunoassay reagent; nutritional supplement; therapeutic activity; ds.
XX
OS Homo sapiens.
XX

```

PN W09746683-A2.
XX
PD 11-DEC-1997.
XX
XX 06-JUN-1997; 97WO-US09878.
XX
PR 07-JUN-1996; 96US-0659224.
XX
PA (GEMV) GENETICS INST INC.
XX
PI Bowman M, Evans C, Jacobs K, Lavallie ER, McCoy JM;
PI Merberg D, Racie LA, Spaulding V, Treacy M;
DR WPI: 1998-042191/04.
XX
PT Nucleic acids encoding secreted proteins from clones within ATCC
PT 98076 - useful as immuno-modulators, anti-proliferative agents,
PT regulators of cell differentiation and tissue growth, etc
XX
PS Claim 13; Page 57; 99pp; English.
XX
CC The present sequence represents the nucleotide sequence of the 3'
CC portion of clone A2302.1. The clone was isolated from a human colon
CC (Caca-2 adenocarcinoma) cDNA library using probe AAV05749. A2302.1 is
CC a full length clone encoding a secreted protein. AAV05719 provides the
CC 5' portion sequence and AAV05718 the internal sequence. The A2302.1
CC sequence shows some homology with an EST sequence. The nucleic acid
CC can be used for expression of recombinant proteins, as tissue, molecular
CC weight or chromosome markers. Indicators of genetic disorders and
CC sources of probes and primers. They can also be used to generate
CC anti-protein or anti-DNA antibodies and as components of interaction trap
CC assays etc. The protein is useful for raising antibodies, as immunoassay
CC reagents and as nutritional supplements. The protein may possibly have
CC any of a great variety of therapeutic activities.
XX
SQ Sequence 85 BP; 80 A; 0 C; 0 G; 5 T; 0 other;

Query Match 1.8%; Score 34; DB 19; Length 85;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
DB 5 taaaaaaaaaaaaaaaaaaaaaaaaaaaaa 38

RESULT 13
AAQ11760/c
ID AAQ11760 standard; DNA; 86 BP.
XX
AC AAQ11760;
XX
DT 22-JUL-1991 (first entry)
XX
DE Self-complementary, T7 promoter hairpin-forming sequence #2.
XX
KM T7 bacteriophage; DNA-dependent RNA polymerase;
KM target sequence amplification; ss.
XX
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT 1..86
FT /tag= a
FT /note= "number of T's in loop is 50"
XX
PN BP427073-A.
XX
PD 15-MAY-1991.
XX
PF 27-OCT-1990; 90EP-0120650.
XX

PR 23-AUG-1990; 90US-0569991.
PR 09-NOV-1989; 89US-0434372.
XX
XX (MOLE-) MOLECULAR DIAGNOSTI.
XX
XX
PI Dattagupta N;
XX
DR WPI: 1991-141638/20.
XX
PT Nucleic acid probe for amplification and detection of target
PT sequence - capable of forming ligatable hairpin structured
PT promoter and transcribing target sequence, is sensitive and
PT useful in medical diagnosis
XX
PS Claim 2; Page 11; 15pp; English.
XX
CC The sequence is an example of a preferred T7 RNA polymerase
CC promoter for use in the invention. There can be from 2 to 50 T's in
CC the loop region. A probe sequence is ligated to the 3' end of the
CC promoter region. Upon hybridisation of the probe to a target sequence
CC and ligation of the hybridised target sequence to the 5' end of the
CC hairpin-forming sequence, the target sequence can be transcribed by
CC T7 RNA polymerase. This allows the target sequence to be amplified
CC (and detected) using a single oligonucleotide component (c.f. PCR).
XX
SQ Sequence 86 BP; 12 A; 6 C; 6 G; 62 T; 0 other;

Query Match 1.8%; Score 34; DB 12; Length 86;
Best Local Similarity 100.0%; Pred. No. 0.0004;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
DB 70 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 37

RESULT 14
AAC78380
ID AAC78380 standard; CDNA; 89 BP.
XX
AC AAC78380;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:774.
XX
XX Human; cancer associated gene; cancer antigen; detection; cancer;
KM diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;
KM antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KM antiinflammatory; antihypertoid; antiallergic; antibacterial; caridiant;
KM dermatologic; neuroprotective; thrombolytic; coagulant; motropic;
KM vasotrophic; antiproliferative; angiogenic; gene therapy; inflammation;
KM immune disorder; haematopoietic cell disorder; autoimmune disorder;
KM allergic reaction; graft versus host disease; organ rejection;
KM haemostatic; thrombolytic; cardiovascular disorder; infection;
KM neurological disease; drug screening; ss.
XX
XX Homo sapiens.
OS
XX
PN W020005350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX

DR WPI: 2000-587533/55.
DR P-PSDB: AAB44171.

XX Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX

PS Claim 1; Page 1321; 2352pp; English.

CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antihypertensive; antiallergic; thrombolytic; antiviral;
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC neurotropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX

SO Sequence 89 BP; 64 A; 4 C; 9 G; 5 T; 7 other;

Query Match 1.8%; Score 34; DB 21; Length 89;
Best Local Similarity 100.0%; Pred. No. 0.0004;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taataaaaaaaaaaaaaaaaaaaaaa 1912
DB 25 taataaaaaaaaaaaaaaaaaaaaaa 58

RESULT 15
AAQ11762/c

ID AAQ11762 standard; DNA; 90 BP.

AC AAQ11762;

DT 22-JUL-1991 (first entry)

DE Self-complementary, T3 promoter hairpin-forming sequence #2.

XX T3 bacteriophage; DNA-dependent RNA polymerase;

KW target sequence amplification; ss.

OS Synthetic.

XX Key Location/Qualifiers

FT stem_loop 1..90

FT /tag= a /note= "number of T's in loop is 50"

PN EP427073-A.

XX 15-MAY-1991.

PF 27-OCT-1990; 90BP-0120650.

XX 23-AUG-1990; 90US-0569991.

PR 09-NOV-1989; 89US-0434372.

XX (MOLE-) MOLECULAR DIAGNOSTI.

XX

PI Datta Gupta N;

XX WPI: 1991-141638/20.

DR Nucleic acid probe for amplification and detection of target
XX sequence - capable of forming ligatable hairpin structured
PT promoter and transcribing target sequence, is sensitive and
PT useful in medical diagnosis

XX Claim 2; Page 11; 15pp; English.

XX The sequence is an example of a preferred T3 RNA polymerase
CC promoter for use in the invention. There can be from 2 to 50 T's in
CC the loop region. A probe sequence is ligated to the 3' end of the
CC promoter region. Upon hybridisation of the probe to a target sequence
CC and ligation of the hybridised target sequence to the 5' end of the
CC hairpin-forming sequence, the target sequence can be transcribed by
CC T3 RNA polymerase. This allows the target sequence to be amplified
CC (and detected) using a single oligonucleotide component (c.f. PCR).
CC See also AAQ11759-011761 and AAQ11763-011764.

XX Sequence 90 BP; 12 A; 8 C; 8 G; 62 T; 0 other;

Query Match 1.8%; Score 34; DB 12; Length 90;
Best Local Similarity 100.0%; Pred. No. 0.0004;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taataaaaaaaaaaaaaaaaaaaaaa 1912
DB 71 TAAAAAATAAAAAAAAAAAAAAAAAAAAAA 38

Search completed: November 13, 2001, 09:56:04
Job time: 4280 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2001, 08:43:39 ; Search time 100.16 Seconds
(without alignments)
4323.342 Million cell updates/sec

Title: US-09-589-510-3
Perfect score: 1912
Sequence: 1 accacagctccgcaattt.....aaaaaaaaaaaaaaaaaa 1912

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 351203 segs, 113238999 residues
Word size : 0
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: /cgn2-6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2-6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2-6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2-6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2-6/ptodata/2/ina/PCBUS.COMB.seq.*
6: /cgn2-6/ptodata/2/ina/Backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	1.8	55	1	US-08-113-646A-41 Sequence 41, Appl
2	34	1.8	114	1	US-08-120-827-99 Sequence 99, Appl
3	34	1.8	114	1	US-08-478-675-99 Sequence 99, Appl
4	34	1.8	120	1	US-08-153-051B-28 Sequence 28, Appl
5	34	1.8	120	1	US-08-060-952C-44 Sequence 44, Appl
6	34	1.8	120	2	US-08-151-477A-28 Sequence 28, Appl
7	34	1.8	120	3	US-08-819-867-58 Sequence 38, Appl
8	34	1.8	140	1	US-08-628-417-5 Sequence 5, Appl
9	34	1.8	141	3	US-08-737-078A-1 Sequence 1, Appl
10	34	1.8	141	5	PCP-US94-04706-1 Sequence 1, Appl
11	34	1.8	144	1	US-08-702-344-26 Sequence 26, Appl
12	34	1.8	216	1	US-08-686-878A-34 Sequence 15, Appl
13	34	1.8	222	4	US-08-481-190-15 Sequence 15, Appl
14	34	1.8	222	5	PCP-US93-00869-15 Sequence 15, Appl
15	34	1.8	240	1	US-08-628-417-6 Sequence 6, Appl
16	34	1.8	289	2	US-08-341-568-3 Sequence 3, Appl
17	34	1.8	289	2	US-08-911-020-3 Sequence 3, Appl
18	34	1.8	340	1	US-08-171-385-27 Sequence 27, Appl
19	34	1.8	340	3	US-08-361-441B-27 Sequence 2, Appl
20	34	1.8	347	1	US-08-104-072B-20 Sequence 14, Appl
21	34	1.8	350	1	US-08-171-385-11 Sequence 11, Appl
22	34	1.8	350	1	US-08-248-016-11 Sequence 11, Appl
23	34	1.8	350	1	US-08-451-501-11 Sequence 11, Appl
24	34	1.8	350	3	US-08-361-441B-14 Sequence 14, Appl
25	34	1.8	350	5	PCP-US95-06761-11 Sequence 11, Appl
26	34	1.8	361	2	US-08-465-380-39 Sequence 39, Appl
27	34	1.8	361	2	US-08-486-397-39 Sequence 39, Appl

28	34	1.8	361	2	US-08-486-399-39 Sequence 39, Appl
29	34	1.8	361	2	US-08-461-965-39 Sequence 39, Appl
30	34	1.8	361	2	US-08-634-641-39 Sequence 39, Appl
31	34	1.8	361	3	US-09-249-471-39 Sequence 39, Appl
32	34	1.8	361	3	US-09-249-472-39 Sequence 39, Appl
33	34	1.8	361	3	US-09-249-451-39 Sequence 39, Appl
34	34	1.8	361	3	US-08-809-455-39 Sequence 39, Appl
35	34	1.8	361	3	US-09-249-461-39 Sequence 39, Appl
36	34	1.8	361	3	US-09-249-448-39 Sequence 23, Appl
37	34	1.8	375	3	US-08-946-026-23 Sequence 13, Appl
38	34	1.8	399	1	US-07-885-970A-13 Sequence 12, Appl
39	34	1.8	399	1	US-08-530-797-12 Sequence 12, Appl
40	34	1.8	399	1	US-08-298-829-13 Sequence 12, Appl
41	34	1.8	399	1	US-08-787-835-12 Sequence 12, Appl
42	34	1.8	399	2	US-08-688-988-44 Sequence 44, Appl
43	34	1.8	444	3	US-08-524-757-1 Sequence 1, Appl
44	34	1.8	458	1	US-08-924-759-7 Sequence 7, Appl
45	34	1.8	458	2	US-08-924-759-7 Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-113-646A-41
Sequence 41, Application US/08113646A
Patent No. 5578468
GENERAL INFORMATION:
APPLICANT: PICKUP, David J.
APPLICANT: PATEL, Dhaval Kumar
APPLICANT: ANTICZAK, James B.
TITLE OF INVENTION: SITE-SPECIFIC RNA CLEAVAGE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESSES:
ADDRESSEE: NIXON & VANDERHIE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/113.646A
FILING DATE: 31-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/084,406
FILING DATE: 10-AUG-1987
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 1579-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 55 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA (genomic)
US-08-113-646A-41

Query Match 1.8%; Score 34; DB 1; Length 55;
Best Local Similarity 97.1%; Pred. No. 16-05; Gaps 0;
Matches 33; Conservative 1; Mismatches 0; Indels 0;

OY 1879 taaaaaaaaaaaaaaaaaaaaa 1912
:|||||
DB 15 UAAAAAAAAAAAAAAAAAAAAA 48

RESULT 2

US-08-120-827-99
; Sequence 99, Application US/08120827
; Patent No. 5525495
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/120,827
; FILING DATE: 15-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5525495man F.
; REGISTRATION NUMBER: 24,618
; TELEPHONE: (703)413-3000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEFAX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 99:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 114 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-08-120-827-99

Query Match 1.8%; Score 34; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 9.7e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taaaaaaaaaaaaaaaaaaaaa 1912
:|||||
DB 80 TAAAAAAAAAAAAAAAAAAAAA 113

RESULT 3

US-08-478-675-99
; Sequence 99, Application US/08478675
; Patent No. 5773246
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,675
; FILING DATE: 07-JUN-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/120,827
; FILING DATE: 15-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5773246man F.
; REGISTRATION NUMBER: 24,618
; TELEPHONE: (703)413-3000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEFAX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 99:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 114 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-08-478-675-99

NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,675
FILING DATE: 07-JUN-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/120,827
FILING DATE: 15-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5773246man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 714-158-0 CIP
TELEPHONE: (703)413-3000
TELEFAX: (703)413-2220
TELEFAX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 114 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-478-675-99

Query Match 1.8%; Score 34; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 9.7e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taaaaaaaaaaaaaaaaaaaaa 1912
:|||||
DB 80 TAAAAAAAAAAAAAAAAAAAAA 113

RESULT 4

US-08-153-051B-28
; Sequence 28, Application US/08153051B
; Patent No. 5645986
; GENERAL INFORMATION:
; APPLICANT: Michael D. West
; APPLICANT: Jerry W. Shay
; APPLICANT: Woodring E. Wright
; APPLICANT: Elizabeth Blackburn
; APPLICANT: Nam Woo Kim
; APPLICANT: Calvin B. Harley
; APPLICANT: Scott L. Weinrich
; APPLICANT: Catherine Strahl
; APPLICANT: Michael J. McEachern
; TITLE OF INVENTION: Homayoun Vaziri
; TITLE OF INVENTION: THERAPY RELATED TO TELOMERE
; TITLE OF INVENTION: CONDITIONS RELATED TO TELOMERE
; TITLE OF INVENTION: LENGTH AND/OR TELOMERASE ACTIVITY
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California

COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSEO Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,051B
FILING DATE: No. 564596ember 12, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/038,766
FILING DATE: March 24, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/195
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 120
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-153-051B-28

Query Match 1.8%; Score 34; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 9.7e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
Db 73 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 106

RESULT 5
US-08-060-952C-44
Sequence 44, Application US/08060952C
Patent No. 5695932
GENERAL INFORMATION:
APPLICANT: Michael D. West
APPLICANT: Jerry W. Shay
APPLICANT: Woodring E. Wright
APPLICANT: Elizabeth Blackburn
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF CONDITIONS
RELATED TO TELOMERE LENGTH AND/OR
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/060,952C
FILING DATE: May 13, 1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/882,438

FILING DATE: May 13, 1992
APPLICATION NUMBER: 08/038,766
FILING DATE: March 24, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 202/045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 120
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-060-952C-44

Query Match 1.8%; Score 34; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 9.7e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1879 taaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
Db 73 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 106

RESULT 6
US-08-151-477A-28
Sequence 28, Application US/08151477A
Patent No. 5830644
GENERAL INFORMATION:
APPLICANT: Michael D. West
APPLICANT: Jerry W. Shay
APPLICANT: Woodring E. Wright
APPLICANT: Elizabeth Blackburn
APPLICANT: Nam Woo Kim
APPLICANT: Calvin B. Harley
APPLICANT: Scott L. Weinrich
APPLICANT: Catherine Strahl
APPLICANT: Michael J. McEachern
APPLICANT: Homayoun Vaziri
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF
CONDITIONS RELATED TO TELOMERE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSEO Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/151,477A
FILING DATE: No. 5830644ember 12, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/038,766
FILING DATE: March 24, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 202/189
TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 120
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-151-477A-28

Query Match
Best Local Similarity 1.8%; Score 34; DB 2; Length 120;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
Db 73 TAAAAAAAAAAAAAAAAAAAAAAAAA 106

RESULT 7
US-08-819-867-58

; Sequence 58, Application US/08819867
; Patent No. 6007989

; GENERAL INFORMATION:

; APPLICANT: Michael D. West

; APPLICANT: Calvin B. Harley

; APPLICANT: Scott L. Weinrich

; APPLICANT: Catherine M. Strahl

; APPLICANT: Michael J. McEachern

; APPLICANT: Jerry Shay

; APPLICANT: Woodring E. Wright

; APPLICANT: Elizabeth H. Blackburn

; APPLICANT: Nam Woo Kim

; APPLICANT: Homayoun Vaziri

; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF

; TITLE OF INVENTION: CONDITIONS RELATED TO

; TITLE OF INVENTION: TELOMERE LENGTH AND/OR

; TITLE OF INVENTION: TELOMERASE ACTIVITY

; NUMBER OF SEQUENCES: 80

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; CITY: Suite 4700

; STATE: California

; COUNTRY: U.S.A.

; ZIP: 90071-2066

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 MB

; MEDIUM TYPE: storage

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: FastSeq for Windows 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/819,867

; FILING DATE: March 14, 1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/153,051

; FILING DATE: No. 6007989ember 12, 1993

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Chambers, Daniel M.

; REGISTRATION NUMBER: 34,561

; REFERENCE/DOCKET NUMBER: 224/232

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:
LENGTH: 120 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-819-867-58

Query Match
Best Local Similarity 1.8%; Score 34; DB 3; Length 120;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
Db 73 TAAAAAAAAAAAAAAAAAAAAAAAAA 106

RESULT 8
US-08-628-417-5

; Sequence 5, Application US/08628417

; Patent No. 5627054

; GENERAL INFORMATION:

; APPLICANT: GILLESPIE, DAVID

; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC

; TITLE OF INVENTION: POLYMERASE CHAIN REACTION

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL

; STREET: DEFENSE COMMAND

; CITY: ABERDEEN PROVING GROUND

; STATE: MARYLAND

; COUNTRY: USA

; ZIP: 21010-5423

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/628,417

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: BIFFONT, ULYSSES J

; REGISTRATION NUMBER: 39,908

; REFERENCE/DOCKET NUMBER: DAM 398-94

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 410-671-1158

; TELEFAX: 410-671-2534

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 140 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: oligodeoxynucleotide

; HYPOTHETICAL: NO

; ANTI-SENSE: YES

; US-08-628-417-5

Query Match
Best Local Similarity 1.8%; Score 34; DB 1; Length 140;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1879 taaaaaaaaaaaaaaaaaaaaaaaaa 1912
|||||
Db 39 TAAAAAAAAAAAAAAAAAAAAAAAAA 72

RESULT 9

US-08-737-078A-1/c

```
Sequence 1, Application US/08737078A
Patent No. 6027934
GENERAL INFORMATION:
APPLICANT: Powell, Curtis
TITLE OF INVENTION: VACCINE FOR, DIAGNOSTIC ASSAY FOR AND
TITLE OF INVENTION: METHOD OF TREATING PARASITIC HEMOFLAGELLATE PROTOZOA
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect 3.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,078A
FILING DATE: 16-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04931
FILING DATE: 11-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/714,464
FILING DATE: 13-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/365,413
FILING DATE: 13-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: POWELL 203-PCT-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-737-078A-1

Query Match 1.8%; Score 34; DB 3; Length 141;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 taaaaaaaaaaaaaaaaaaaaaaaaa 1912
DB 61 TAAAAAAAAAAAAAAAAAAAAAAAAA 28

RESULT 10
PCT-US94-04706-1/C
Sequence 1, Application PC/TUS9404706
GENERAL INFORMATION:
APPLICANT: Powell, Curtis
TITLE OF INVENTION: VACCINE FOR, DIAGNOSTIC ASSAY FOR AND
TITLE OF INVENTION: METHOD OF TREATING PARASITIC
TITLE OF INVENTION: HEMOFLAGELLATE PROTOZOA
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Horn Kramer & Woods
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.0
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04706
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04931
FILING DATE: 11-JUN-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/714,464
FILING DATE: 13-JUN-91
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/365,413
FILING DATE: 13-JUN-89
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: POWELL 203-PCT-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US94-04706-1
```

```
Query Match 1.8%; Score 34; DB 5; Length 141;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 taaaaaaaaaaaaaaaaaaaaaaaaa 1912
DB 61 TAAAAAAAAAAAAAAAAAAAAAAAAA 28

RESULT 11
US-08-702-344-26
Sequence 26, Application US/08702344
Patent No. 5723315
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavalie, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702,344
FILING DATE:
```

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-702-344-26

Query Match 1.8%; Score 34; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 9.5e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 taataataataataataataataataataata 1912
|||||
Db 22 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55

RESULT 12
US-08-686-878A-34
Sequence 34, Application US/08686878A
Patent No. 5708157
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallie, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Evans, Cheryl
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/686,878A
APPLICATION NUMBER: US/08/686,878A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 216 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-686-878A-34

Query Match 1.8%; Score 34; DB 1; Length 216;
Best Local Similarity 100.0%; Pred. No. 9.3e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 taataataataataataataataataataata 1912
|||||
Db 168 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 201

RESULT 13
US-08-481-190-15
Sequence 15, Application US/08481190
Patent No. 6160204
GENERAL INFORMATION:
APPLICANT: John C. Steffens
TITLE OF INVENTION: Polyphehol Oxidase CDNA
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Yahwak & Associates
STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,190
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 203,533
FILING DATE: 02-24-1994
ATTORNEY/AGENT INFORMATION:
NAME: George M. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: UA 816 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)268-1951
TELEFAX: (203)268-1951
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 222 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-481-190-15

Query Match 1.8%; Score 34; DB 4; Length 222;
Best Local Similarity 100.0%; Pred. No. 9.3e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 taataataataataataataataataataata 1912
|||||
Db 176 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 209

RESULT 14
PCT-US93-00869-15
Sequence 15, Application PC/TUS9300869
GENERAL INFORMATION:
APPLICANT: John C. Steffens
TITLE OF INVENTION: Polyphehol Oxidase CDNA: Cloning
TITLE OF INVENTION: and Applications
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Yahwak & Associates

```

: STREET: 25 Skytop Drive
: CITY: Trumbull
: STATE: Connecticut
: COUNTRY: USA
: ZIP: 06611
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: Macintosh
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Microsoft Word 4.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/00869
: FILING DATE: 19930129
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: George M. Yahwak
: REGISTRATION NUMBER: 26,824
: REFERENCE/DOCKET NUMBER: CRF D-1057
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (203)268-1951
: TELEFAX: (203)268-1951
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 222 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: PCT-US93-00869-15

```

```

Query Match          1.8%; Score 34; DB 5; Length 222;
Best Local Similarity 100.0%; Pred. No. 9.3e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 taataaaaaaaaaaaaaaaaaaaaaaaaaa 1912
      ||||||||||||||||||||||||||||
Db 176 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 209

```

```

RESULT 15
US-08-628-417-6
: Sequence 6, Application US/08628417
: Patent No. 5627054
: GENERAL INFORMATION:
: APPLICANT: GILLESPIE, DAVID
: TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
: ADDRESS: DEFENSE COMMAND
: STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
: CITY: ABERDEEN PROVING GROUND
: STATE: MARYLAND
: COUNTRY: USA
: ZIP: 21010-5423
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/628,417
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: BIFFONT, ULYSSES J
: REGISTRATION NUMBER: 39,908
: REFERENCE/DOCKET NUMBER: DAM 398-94
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 410-671-1158
: TELEFAX: 410-671-2534

```

```

: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 240 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: oligodeoxynucleotide
: HYPOTHEICAL: NO
: ANTI-SENSE: YES
: US-08-628-417-6

```

```

Query Match          1.8%; Score 34; DB 1; Length 240;
Best Local Similarity 100.0%; Pred. No. 9.2e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1879 taataaaaaaaaaaaaaaaaaaaaaaaaaa 1912
      ||||||||||||||||||||||||||||
Db 39 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 72

```

Search completed: November 13, 2001, 09:53:18
Job time: 4179 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2001, 07:53:44 ; Search time 1462.74 Seconds
(without alignments)
12356.164 Million cell updates/sec

```

Title: US-09-589-510-3
Perfect score: 1912
Sequence: 1 acccagcgttcgcaatt.....aaaaaaaaaaaaaaaa 1912

```

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1022815 seqs, 4726426750 residues

word size : 0

Total number of hits satisfying chosen parameters: 20456230

Minimum	DB	seq	length:	0
---------	----	-----	---------	---

Maximum DB seq length: 20000000000

Post-processing: Listing first 45 summaries

Database :

EST: *

1:	qb_est1:*
2:	qb_est2:*
3:	qb_est3:*
4:	qb_est4:*
5:	qb_est5:*
6:	qb_est6:*
7:	qb_est7:*
8:	qb_est8:*
9:	qb_est9:*
10:	qb_est10:*
11:	qb_est11:*
12:	qb_est12:*
13:	qb_est13:*
14:	qb_est14:*
15:	qb_est15:*
16:	qb_est16:*
17:	qb_est17:*
18:	qb_est18:*
19:	qb_est19:*
20:	qb_est20:*
21:	qb_est21:*
22:	qb_est22:*
23:	qb_est23:*
24:	qb_est24:*
25:	qb_est25:*
26:	qb_est26:*
27:	qb_est27:*
28:	qb_est28:*
29:	qb_est29:*
30:	qb_est30:*
31:	qb_est31:*
32:	qb_est32:*
33:	qb_est33:*
34:	qb_est34:*
35:	qb_est35:*
36:	qb_est36:*
37:	qb_est37:*
38:	qb_est38:*
39:	qb_est39:*
40:	qb_est40:*
41:	qb_est41:*
42:	qb_est42:*
43:	qb_est43:*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: em_gss_fun1:*
182: em_gss_fun2:*
183: em_gss_fun3:*
184: em_gss_fun4:*
185: em_gss_fun5:*
186: em_gss_fun6:*
187: em_gss_fun7:*
188: em_gss_fun8:*
189: em_gss_fun9:*
190: em_gss_inv1:*
191: em_gss_inv2:*
192: em_gss_inv3:*
193: em_gss_inv4:*
194: em_gss_inv5:*
195: em_gss_inv6:*
196: em_gss_inv7:*
197: em_gss_inv8:*
198: em_gss_inv9:*
199: em_gss_inv10:*
200: em_gss_inv11:*
201: em_gss_inv12:*
202: em_gss_inv13:*
203: em_gss_inv14:*
204: em_gss_inv15:*
205: em_gss_inv16:*
206: em_gss_inv17:*
207: em_gss_inv18:*
208: em_gss_inv19:*
209: em_gss_inv20:*
210: em_gss_inv21:*
211: em_gss_inv22:*
212: em_gss_inv23:*
213: em_gss_inv24:*
214: em_gss_inv25:*
215: em_gss_inv26:*
216: em_gss_inv27:*
217: em_gss_inv28:*
218: em_gss_inv29:*
219: em_gss_inv30:*
220: em_gss_inv31:*
221: em_gss_inv32:*
222: em_gss_inv33:*
223: em_gss_inv34:*
224: em_gss_inv35:*
225: em_gss_inv36:*
226: em_gss_inv37:*
227: em_gss_inv38:*
228: em_gss_inv39:*
229: em_gss_inv40:*
230: em_gss_inv41:*
231: em_gss_inv42:*
232: em_gss_inv43:*
233: em_gss_inv44:*
234: em_gss_inv45:*
235: em_gss_inv46:*
236: em_gss_inv47:*
237: em_gss_inv48:*
238: em_gss_inv49:*
239: em_gss_inv50:*
240: em_gss_inv51:*
241: em_gss_inv52:*
242: em_gss_inv53:*
243: em_gss_inv54:*
244: em_gss_inv55:*
245: em_gss_inv56:*
246: em_gss_inv57:*
247: em_gss_inv58:*
248: em_gss_inv59:*
249: em_gss_inv60:*
250: em_gss_inv61:*
251: em_gss_inv62:*
252: em_gss_inv63:*
253: em_gss_inv64:*
254: em_gss_inv65:*
255: em_gss_inv66:*
256: em_gss_inv67:*
257: em_gss_inv68:*
258: em_gss_inv69:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

QY 579 gactgttaaggagactgaacacgaagtagattcttcaattatgatgtctgacaa 638
|||||
Db 181 GACTGTTAAAGGAGCTAGCAAGTAAGATTGATTTATGATGCTGATCAA 240
QY 639 ggaagaag 646
|||||
Db 241 GGAAGAAG 248

RESULT 4
BE512425/c 549 bp mRNA EST 07-AUG-2000
LOCUS 946071A06.x1 946 - tassal primordium prepared by Schmidt lab Zea
DEFINITION
mays cDNA, mRNA sequence.
ACCESSION BE512425
VERSION BE512425.1 GI:9733673
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 549)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 946071 row: A column: 06.
location/Qualifiers
1. 549
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassal primordium prepared by Schmidt
lab"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="XLOLR"
/note="Organ: tassels; Vector: HybridZAP; Site: 1; EcoRI;
Site: 2; XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridZAP. Sample insert size range was 350 bp
to 3 Kb with a 1 Kb average."
BASE COUNT 150 a 146 c 116 g 137 t
ORIGIN

Query Match 7.0%; Score 134; DB 136; Length 549;
Best Local Similarity 99.0%; Pred. No. 8.4e-57;
Matches 284; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1240 attgtatgatgaagaagctcttctatttaagcgaagatcgacacagacatcttg 1299
|||||
Db 519 ATTGATATGATGAGAAAGATTCTTCTATTATTAGCGAGATCGACACAGACATCTTTA 460
QY 1300 agacatgcatcaattatgatactaccgcagcgtgctcacaagaactaaggaagag 1359
|||||
Db 459 AGACATGCTATTCATTTGATATCACCCTGCGACGCTGCTCAAAAGACTAATGGAAGAGAG 400
QY 1360 aaaaatcgaagagctgcatcgaaggaagtcagtggtctctatttggatgccaatccctcg 1419
|||||
Db 399 AAAAATATGCAAGCTGATCTCGAGGAAGTCAGTGGGCTCTATTGATGCCAAATCCTCG 340
QY 1420 gctcgcgtctccagagacaagaatacatcacactatgtgatactcctgctgt 1479
|||||

Db 339 GCTCGGCTGCTCCAGGAGCAACAGAAGATACATCACCCTAGATTGGATCCTGCTGT 280
QY 1480 ggaagctcgaagaagaatgtagttgcacactcgaaagatcatctagtg 1526
|||||
Db 279 GGAAGTCTCGAAGAGCAATGATTGCCAGCTCGAAAGTCACTAGTG 233

RESULT 5
BE640273/c 375 bp mRNA EST 30-AUG-2000
LOCUS 946082B06.x1 946 - tassal primordium prepared by Schmidt lab Zea
DEFINITION
mays cDNA, mRNA sequence.
ACCESSION BE640273
VERSION BE640273.1 GI:9953690
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 375)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 946082 row: B column: 06.
location/Qualifiers
1. 375
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassal primordium prepared by Schmidt
lab"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="XLOLR"
/note="Organ: tassels; Vector: HybridZAP; Site: 1; EcoRI;
Site: 2; XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridZAP. Sample insert size range was 350 bp
to 3 Kb with a 1 Kb average."
BASE COUNT 106 a 106 c 79 g 84 t
ORIGIN

Query Match 5.8%; Score 110; DB 138; Length 375;
Best Local Similarity 99.4%; Pred. No. 1.3e-44;
Matches 160; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1366 tgcgaagctcatctcgagaagtcagtggtctatttgaatgccaatcctcgctcg 1425
|||||
Db 364 TGCAGAGCTGATCTCGAGGAAGTCAAGTGGGCTCTATTGATGCCAAATCCTCGGCTCG 305
QY 1426 ctgttcagagagaacaagaatatcacactagattggatcctctgctggaagt 1485
|||||
Db 304 CTGCTCAGGAGCAACAGAAGATACATCACTAGATTGGATCAGCTGCTGGAAGT 245
QY 1486 ctgcgaagaatgtagttgcacactcgaaagatcatctagtg 1526
|||||
Db 244 CTCGAAGAGATGTAGTTGCCAGCTCGAAAGTCACTAGTG 204

RESULT 6
BG316941/c 109 bp mRNA EST 26-FEB-2001
LOCUS BG316941

DEFINITION 947024C05.x2 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA sequence.

ACCESSION BG316941

VERSION BG316941.1 GI:13126371

KEYWORDS EST

SOURCE Zea mays.

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 109)

AUTHORS Walbot, V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University

JOURNAL Unpublished (1999)

COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 947024 row: C column: 05.

FEATURES

source

1..109
/location/Qualifiers

/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="947 - 2 week shoot from Barkan lab"
/tissue_type="leaf and stem, including leaf base"
/dev_stage="2 week old seedling (3 leaves)"
/lab_host="X11-Blue"
/note="Organ: shoot; Vector: lambda ZAP (pBluescript SK-); Site:1: EcoRI; Site:2: XhoI; directionally cloned using StrataGene's Unizap XR cDNA cloning kit with the 5' end at the EcoRI site. The library represents 8 x 10⁵ independent recombinant phage. The plants were greenhouse grown."

BASE COUNT 30 a 29 c 17 g 33 t

ORIGIN

Query Match 5.7%; Score 109; DB 152; Length 109;
Best Local Similarity 100.0%; Pred. No. 4.6e-44;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1214 tggctatcgagcagacaagtgtgagagattgatgatgaaagaattcttatttag 1273
|||||

Db 109 tggctatcgagcagacaagtgtgagagattgatgatgaaagaattcttatttag 50
|||||

QY 1274 ggcagatcgagcagacacatcttgcagacatgcattcatcattgattac 1322
|||||

Db 49 ggcagatcgagcagacacatcttgcagacatgcattcatcattgattac 1
|||||

RESULT 7

LOCUS AW927557 512 bp mRNA EST 30-MAY-2000

DEFINITION 945018G05.X2 945 - Mixed adult tissues from Walbot lab, same as 707 (SK) Zea mays cDNA, mRNA sequence.

ACCESSION AW927557

VERSION AW927557.1 GI:8102879

KEYWORDS EST

SOURCE Zea mays.

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 512)

AUTHORS Walbot, V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University

JOURNAL Unpublished (1999)

COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 945018 row: G column: 05.

FEATURES

source

1..555
/location/Qualifiers

/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="945 - Mixed adult tissues from Walbot lab,"

COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 945018 row: G column: 05.

FEATURES

source

1..512
/location/Qualifiers

/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="945 - Mixed adult tissues from Walbot lab, same as 707 (SK)"
/tissue_type="tassel, kernel, silk, husk, root, leaf"
/dev_stage="fully-grown"
/lab_host="DH10B"
/note="Organ: tassel, kernel, silk, husk, root, leaf; Vector: pGAD10; Site:1: EcoRI; cDNA library from fully differentiated maize tissues from an active mutator plant. Tissue ratio is 4:2:1:1:1 (tassel, kernel, silk, husk, root, leaf). Unidirectionally cloned. New library number given to library 707 for additional sequencing."

BASE COUNT 148 a 98 c 151 g 115 t

ORIGIN

Query Match 4.9%; Score 94; DB 122; Length 512;
Best Local Similarity 100.0%; Pred. No. 1.8e-36;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 gaggtcaagaagaactggggtgctatgaaatttcgtagacttaagttgctata 465
|||||

Db 177 gaggtcaagaagaactggggtgctatgaaatttcgtagacttaagttgctata 236
|||||

QY 466 aaggaacaacaagaaggtttatgaagagaagttta 499
|||||

Db 237 AAGGAAAACAAGAGGTTTATGAAAGGAGAGCTTA 270
|||||

RESULT 8

LOCUS AW927556 555 bp mRNA EST 30-MAY-2000

DEFINITION 945018G05.X1 945 - Mixed adult tissues from Walbot lab, same as 707 (SK) Zea mays cDNA, mRNA sequence.

ACCESSION AW927556

VERSION AW927556.1 GI:8102878

KEYWORDS EST

SOURCE Zea mays.

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 555)

AUTHORS Walbot, V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University

JOURNAL Unpublished (1999)

COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 945018 row: G column: 05.

FEATURES

source

1..555
/location/Qualifiers

/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="945 - Mixed adult tissues from Walbot lab,"

same as 707 (SK)"
 /tissue_type="tassel, kernal, silk, husk, root, leaf"
 /dev_stage="fully-grown"
 /lab_host="PH10B"
 /note="Organ: tassel, kernal, silk, husk, root, leaf;
 vector: PGAD10; Site_1: EcorI; cDNA library from fully
 differentiated maize tissues from an active Mutator plant.
 tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk,
 root, leaf). Undirectionally cloned. New library number
 given to library 707 for additional sequencing."
 BASE COUNT 156 a 111 c 165 g 123 t
 ORIGIN

Query Match 4.9%; Score 94; DB 122; Length 555;
 Best Local Similarity 100.0%; Pred. No. 1.8e-36;
 Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 gaagtcagaagaactgagtcgactgagaaattccgtagagctatagttgcgata 465
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 194 GAGGTCAAGAAAGTGAAGTGTGATGGAATAATTCCTAGACTATAGTTGGCTATA 253
 QY 466 aaggaagaacaaagagttatgaagaagaggtta 499
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 254 AAGGAAACAAAGAGGTTATGAAGAGAGGTTA 287

RESULT 9
 BG316942 169 bp mRNA EST 26-FEB-2001
 LOCUS 947024C05.y1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA
 DEFINITION sequence.
 ACCESSION BG316942
 VERSION BG316942.1 GI:13126372
 KEYWORDS EST.

SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 169)
 Walbot.V.

REFERENCE
 AUTHORS Maize ESTs from various cDNA libraries sequenced at Stanford
 TITLE University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 947024 row: C column: 05.

FEATURES
 source
 1. 169
 Location/Qualifiers
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone_lib="947" 2 week shoot from Barkan lab"
 /tissue_type="leaf and stem, including leaf base"
 /dev_stage="2 week old seedling (3 leaves)"
 /lab_host="XL1-Blue"
 /note="Organ: shoot; Vector: Lambda ZAP (pBluescript SK-);
 Site_1: EcorI; Site_2: XhoI; Directionally cloned using
 Stratagene's Unizap XR cDNA cloning kit with the 5' end
 at the EcorI site. The library represents 8 x 10e5
 independent recombinant phage. The plants were greenhouse
 grown."

BASE COUNT 56 a 27 c 36 g 50 t
 ORIGIN

Query Match 4.6%; Score 88; DB 152; Length 169;

Best Local Similarity 100.0%; Pred. No. 2.1e-33;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 883 atgaaccacgaagaactcgaatccgcaaaatactgcccagaataataagtgta 942
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 ATGAACACACGAAAGACGAAATCACCGAAACACCAAAATTAATTAAGGTGCTA 60
 QY 943 aatagatatcagatgaggaattgcag 970
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 AATAGATATATGATGATGAGGAAATTCAG 88

RESULT 10
 BG464307 535 bp mRNA EST 20-MAR-2001
 LOCUS EM1_72.B07.B1.A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
 DEFINITION sequence.
 ACCESSION BG464307
 VERSION BG464307.1 GI:13392679
 KEYWORDS EST.
 SOURCE sorghum.

ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 535)
 Reid S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
 An EST database from Sorghum: developing embryos
 Unpublished (2000)
 CONTACT: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu

SEQUENCES have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 517
 POLYA=No.

FEATURES
 source
 1. 535
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Embryo 1 (EM1)"
 /note="Organ: Embryos germinated for 24 hr; Vector:
 pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
 EcorI; The library was made from poly-A RNA in the cloning
 vector lambda ZAP II. Clones to be sequenced were
 prepared by mass excision."
 BASE COUNT 127 a 126 c 183 g 99 t
 ORIGIN

Query Match 4.3%; Score 82; DB 154; Length 535;
 Best Local Similarity 99.2%; Pred. No. 2.2e-30;
 Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 341 tagccagaagctcgagcgaaggtcccttctgctcctatgtagtagatcagaagtact 400
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 293 TAGCCAGAGACTGGCGAGCAAGTCCCTTCTGTGCTATGATGATCAAGATGACT 352

QY 401 cctcggaagtcagaagaactcgaatgctgtagtgaatcttcgtagagctataggttc 460
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 353 CCTCAGAGTCAAGAAACATGAGGTGTGATGAAATTCCTAGAGCTATAGTTTGC 412

QY 461 gtataaaggaana 473
 ||||||||||||||||
 Db 413 GTATTAAGGAAAA 425

RESULT 11
 BE512214/c 519 bp mRNA EST 07-AUG-2000
 LOCUS 946066A01.x1 946 - tassal primordium prepared by Schmidt lab Zea
 DEFINITION mays cDNA, mRNA sequence.
 ACCESSION BE512214
 VERSION BE512214.1 GI:9733357
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade: Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 519)
 Walbot,V.
 Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 946066 row: A column: 01.

FEATURES
 source
 1..519
 /organism="Zea mays"
 /cultivar="OH43"
 /db_xref="taxon:4577"
 /clone_lib="946 - tassal primordium prepared by Schmidt
 lab"
 /issue_type="tassels"
 /dev_stage="just after the transition from vegetative to
 inflorescence development"
 /lab_host="X10LR"
 /note="Organ: tassels; Vector: HybriZAP; Site:1: EcorI;
 Site:2: XhoI; George Chuck dissected immature tassels
 between 1mm and 3mm. Sharon Stanfield prepared the cDNA
 library in HybriZAP. Sample insert size range was 350 bp
 to 3 kb with a 1 kb average."

BASE COUNT 155 a 139 c 94 g 131 t
 ORIGIN

Query Match 4.1%; Score 78; DB 136; Length 519;
 Best Local Similarity 99.2%; Pred. No. 2.4e-28;
 Matches 128; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 601 ctgaagtagatcttcattatgatctgcgtcaagaaagtgagcagtgatg 660
 |||||||
 DB 473 CTGAAGTTAGATTCTTCATTATGATGCTCGATCAAGAAAGTGCGCATG 414
 |||||||
 QY 661 gttatatacatcgaagcaaatagtgagcaagtgaaagatgtgatagtatcttt 720
 |||||||
 DB 413 GTTATATACATTGAAGCAAAATGTCGACAGTGAAGAAGTGATGATCTTTT 354
 |||||||
 QY 721 gctacaaga 729
 |||||||
 DB 353 GCTACAGAA 345

RESULT 12
 BG464635 717 bp mRNA EST 20-MAR-2001
 LOCUS EML_72.B07.g1_A002 Embryo 1 (EML) Sorghum bicolor cDNA, mRNA
 DEFINITION sequence.
 ACCESSION BG464635
 VERSION BG464635.1 GI:13393339
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor

REFERENCE
 AUTHORS Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
 TITLE An EST database from Sorghum: developing embryos
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: PolyTmX
 High quality sequence start: 23
 High quality sequence stop: 680
 POLYA=No.

FEATURES
 source
 1..717
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Embryo 1 (EML)"
 /note="Organ: Embryos germinated for 24 hr; Vector:
 pBluescript II from Lambda Zap II; Site:1: XhoI; Site:2:
 EcorI; The library was made from polyA RNA in the cloning
 vector Lambda Zap II. Clones to be sequenced were
 prepared by mass excision."

BASE COUNT 197 a 127 c 182 g 210 t
 ORIGIN

Query Match 3.6%; Score 68; DB 154; Length 717;
 Best Local Similarity 100.0%; Pred. No. 2.8e-23;
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 958 gaagaatcgaagctgacgtgctgtgtgttgcattatgagtgccacatgtgat 1017
 |||||||
 DB 14 GAAGGAATGCGAGCGCTGTGACCTGGTGTTCATTGATGAGTCCACATGTGGAT 73
 |||||||
 QY 1018 atcgaaatg 1025
 |||||||
 DB 74 ATCGAATG 81

RESULT 13
 A1932123 435 bp mRNA EST 02-AUG-1999
 LOCUS 618021C05.x1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
 DEFINITION sequence.
 ACCESSION A1932123
 VERSION A1932123.1 GI:5670860
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade: Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 435)
 Walbot,V.
 Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu

Plate: 618021 row: C column: 05.
 FEATURES
 Location/Qualifiers

1. .435

/organism="Zea mays"

/cultivar="Ohio43"

/db_xref="taxon:4577"

/clone_lib="618 - Inbred Tassel cDNA Library"

/tissue_type="tassel"

/dev_stage="tassel length from 0.1 to 2.5 cm"

/lab_host="X10LR"

/note="organ: tassel; Vector: PAD-GAL4-2.1 (Hybridap); Inbred tassel library from Schmidt lab"

BASE COUNT 109 a 114 c 92 g 120 t

ORIGIN

Query Match 3.4%; Score 65; DB 103; Length 435;
 Best Local Similarity 100.0%; Pred. No. 9.7e-22;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1249 gatgaagaagctgtctattatgaagcagatcgacagacacatctttgagacatgct 1308
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 404 GATGAAGAAGCTGTCTATTATTAGCGAGATCGACACGACATCTTTGAGACATGCT 345

Oy 1309 atca 1313
 ||||||

Db 344 ATTCA 340

RESULT 14

LOCUS

BC049561 465 bp mRNA EST 25-JAN-2001
 DEFINITION EM1_5.G05.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
 sequence.

ACCESSION

BC049561
 BC049561.1 GI:12501394

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
 An EST database from Sorghum: developing embryos
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu

Seq primer: T7

High quality sequence start: 73

High quality sequence stop: 450

POLYA=Yes.

FEATURES

source

1. .465

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Embryo 1 (EM1)"

/note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 116 a 96 c 120 g 133 t

ORIGIN

Query Match 3.4%; Score 65; DB 117; Length 465;
 Best Local Similarity 100.0%; Pred. No. 9.7e-22;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1249 gatgaagaagctgtctattatgaagcagatcgacagacacatctttgagacatgct 1308
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 29 GATGAAGAAGCTGTCTATTATTAGCGAGATCGACACGACATCTTTGAGACATGCT 88

Oy 1309 atca 1313
 ||||||

Db 89 ATTCA 93

RESULT 15

AM566092 503 bp mRNA EST 10-MAR-2000
 LOCUS 660061G12.Y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
 DEFINITION mRNA sequence.

ACCESSION AM566092.1 GI:7227451

VERSION AM566092.1

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 1 (bases 1 to 503)

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 University

JOURNAL Unpublished (1999)

COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 660061 row: G column: 12.

FEATURES
 source

1. .503

/organism="Zea mays"

/cultivar="Ohio43"

/db_xref="taxon:4577"

/clone_lib="660 - Mixed stages of anther and pollen"

/tissue_type="whole premeiotic anthers to pollen shed"

/dev_stage="premeiotic anthers to pollen shed"

/lab_host="X10LR"

/note="Organ: anthers; Vector: lambda Zap; Site 1: EcoRI; Site 2: XhoI; Anther and pollen cDNA library. Directionally sequenced with 5' end at the EcoRI site. Created by Amie Franklin."

BASE COUNT 146 a 103 c 135 g 119 t

ORIGIN

Query Match 3.4%; Score 65; DB 117; Length 503;
 Best Local Similarity 100.0%; Pred. No. 9.6e-22;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1249 gatgaagaagctgtctattatgaagcagatcgacagacacatctttgagacatgct 1308
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 144 GATGAAGAAGCTGTCTATTATTAGCGAGATCGACACGACATCTTTGAGACATGCT 203

Oy 1309 atca 1313
 ||||||

Db 204 ATTCA 208

Wed Nov 14 08:34:07 2001

us-09-589-510-3.oli.rst

Page 10

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 06:55:54 ; Search time 23.2 Seconds
(without alignments)
1188,962 Million cell updates/sec

Title: US-09-589-510-4

Perfect score: 2263
Sequence: 1 MRFEVQSTSKKQRIATHH.....YLDKSSARLLQROEKYIT 455

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :

A.GeneSeq_0601:*

- 1: /SIDS1/gcgdata/geneSeq/geneSeq/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneSeq/geneSeq/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneSeq/geneSeq/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneSeq/geneSeq/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneSeq/geneSeq/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneSeq/geneSeq/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneSeq/geneSeq/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneSeq/geneSeq/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneSeq/geneSeq/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneSeq/geneSeq/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneSeq/geneSeq/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneSeq/geneSeq/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneSeq/geneSeq/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneSeq/geneSeq/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneSeq/geneSeq/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneSeq/geneSeq/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneSeq/geneSeq/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneSeq/geneSeq/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneSeq/geneSeq/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneSeq/geneSeq/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneSeq/geneSeq/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneSeq/geneSeq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2263	100.0	455	22	AAV72563
2	2253	99.6	455	22	AAV72562
3	2222	98.2	455	22	AAV72564
4	2211.5	97.7	456	22	AAV72565
5	2157.5	95.3	438	22	AAV72566
6	1994	88.1	458	21	AAAG20673
7	1787	79.0	401	21	AAAG20674
8	1761	77.8	395	21	AAAG20675
9	1723.5	76.2	456	20	AAW74417
10	1722.5	76.1	456	20	AAW74416
11	1721.5	76.1	456	20	AAW93945

12	1541.5	68.1	418	20	AAV73986
13	973	43.0	463	21	AAAB21093
14	932.5	41.2	471	21	AAAB21094
15	630.5	27.9	334	22	AAAB80611
16	538.5	23.8	242	20	AAAY04324
17	494.5	21.9	128	21	AAAB27851
18	485	21.4	119	21	AAAB27847
19	481	21.3	119	21	AAAB27848
20	477	21.1	121	21	AAAB27849
21	470	20.8	121	21	AAAB27850
22	446	19.7	127	21	AAAB27852
23	434.5	19.2	263	20	AAAB8630
24	364	16.1	85	20	AAAY04309
25	258	11.4	85	21	AAAG02092
26	231	10.2	106	21	AAAG04033
27	135.5	6.0	603	21	AAAG29815
28	135.5	6.0	627	21	AAAG29814
29	132.5	5.9	826	21	AAAY5050
30	127.5	5.6	555	22	AAAB31943
31	126	5.6	2039	19	AAAB56322
32	125	5.5	566	20	AAAY9068
33	125	5.5	566	22	AAAB31932
34	121.5	5.4	391	21	AAAY5245
35	121.5	5.4	391	21	AAAY5246
36	120.5	5.3	391	21	AAAY5244
37	120.5	5.3	644	21	AAAY08320
38	119.5	5.3	444	20	AAAY08327
39	118.5	5.2	918	20	AAAY35654
40	117.5	5.2	440	17	AAAB4600
41	117	5.2	806	21	AAAB05879
42	117	5.2	806	22	AAAB59399
43	115.5	5.1	632	18	AAAB5684
44	115.5	5.1	680	21	AAAG42129
45	115.5	5.1	821	21	AAAG42128

ALIGNMENTS

RESULT	1	
AAV72563		
ID	AAV72563	standard; Protein: 455 AA.
XX		
AC	AAV72563;	
XX		
DT	02-MAY-2001	(first entry)
XX		
DE	Maize RuvB orthologue protein #2.	
XX		
KW	Maize; RuvB orthologue; branch migration; heteroduplex extension; homologous recombination; transformation; transgenic plant.	
XX		
OS	Zea mays.	
XX		
FH	Key	Location/Qualifiers
FT	Region	/note= "High homology region"
FT	Region	214..294
FT	Region	/note= "Putative heptad repeat region"
FT	Region	297..439
FT	Region	/note= "High homology region"
FT	Binding-site	297..305
FT		/label= "Walker Box B
FT		/note= "ATP binding motif"
XX		
XX	WO200105975-A1.	
XX		
XX	25-JAN-2001.	
XX		
XX	13-JUN-2000; 2000WO-US16271.	
XX		
XX	16-JUL-1999; 99US-0144112.	
XX		

Human prostate tum
Human HEL50 DNA he
Yeast HEL50 DNA he
Environmental stre
Fragment of human
Sequence homology
Sequence homology
Protein fragment e
Sequence homology
Protein fragment e
Protein fragment e
Secreted protein e
Human secreted pro
Human secreted pro
Human secreted pro
Arabidopsis thalia
Arabidopsis thalia
Candida albicans p
Amino acid sequenc
Haemophilus paraga
Amino acid sequenc
Neisseria meningit
Neisseria meningit
Neisseria gonorrhoe
E. coli F18H prote
Human CBP1E10 pro
Amino acid sequenc
S4 protein. Homo
Human transnational
Protein tyrosine p
H. pylori ORF 02ge
Arabidopsis thalia
Arabidopsis thalia

PA (PION-) PIONEER HI-BRED INT INC.

XX Mahajan PB:

XX WPI: 2001-159537/16.

DR N-PSDB: AAD02567.

PT Novel maize RuVB nucleic acid useful for modulating levels of maize
PT RuVB in plants, as probes or amplification primers in the detection,
PT quantitation or isolation of gene transcripts -

PS Claim 12: Page 62-63; 87pp: English.

XX The present sequence is a Zea mays RuVB orthologue protein #2. RuVB
CC along with RuVA catalyses the branch migration process, also known as
CC heteroduplex extension, in homologous recombination. RuVB is used for
CC the control of homologous recombination or transformation efficiency in
CC transgenic plants. The RuVB nucleotide may be used as probes or
CC amplification primers for detecting, quantifying or isolating gene
CC transcripts, in detecting deficiencies in the mRNA level during screening
CC for desired transgenic plants, for detecting gene mutations, for
CC monitoring upregulation of expression or changes in enzyme activity, for
CC detecting any number of allelic variants, orthologues or paralogues of
CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
CC also be used for recombinant expression of its encoded polypeptide, or
CC for use as immunogen in preparing and/or screening of antibodies, and in
CC sense or antisense suppression of one or more genes in a host cell,
CC tissue or plant. The RuVB proteins may be used in assays to agonise or
CC antagonise the enzyme function, or as immunogens or antigens for
CC screening antibodies.

XX Sequence 455 AA:

Query Match 100.0%; Score 2263; DB 22; Length 455;

Best Local Similarity 100.0%; Pred. No. 2.8e-190;

Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRTEVOSTSKKORATHTTHIKGLIDANGMAIALAGVGQAAREAGLAVDMIRQK 60
DB 1 mlreevgstskkgrlathchikgldangmalaalagvgqaareaglavdmirqk 60
QY 61 MAGRAVLGAPRATGKTALGIAOELGSKVPFCPMVGSSEVSSVKKTEVLMEMFRRAI 120
DB 61 megRAVLGAPRATGKTALGIAOELGSKVPFCPMVGSSEVSSVKKTEVLMEMFRRAI 120
QY 121 GLRIKENKEVYEGEYTELSPBEAESTTGGAYSISHVITSLKTVGKOLKIDSSITDAL 180
DB 121 glrIkenkeVgeYtelSPbeaesttgGaYSishvITslKtvGkOLkIDSSITdal 180
QY 181 IREKVAAGDVITYEANSAGVKKRGCDSFATYDLEAEYVPDPKGEVHKKEIVQDVTL 240
DB 181 iRekvaagdvITyEanSagvKKrGCdSFATyDLEAEyVPDPKGEvHKKEIvQdVtL 240
QY 241 HQLDANQPOGODILSLMGOMMRKTEITETKROELNKVYNYIDEGIAELVPGULF 300
DB 241 hQldanqPogODiLslMGomMRkTEITETkROELnKVYnYIDeGIAELvPGulF 300
QY 301 IDEVHMLDIECFSYLNRALESPLSPVILATNRGICNVRGDMTSPHGIPVDLDRLYTI 360
DB 301 iDeVhMLdIEcFSyLnRAleSPlSPvILATnRGICnVRGDMtSPHGIPvDLDRlyTI 360
QY 361 RTEYVGPTEMIQILAIRAOVEIDMEESLAVGEIGQOTSURHAIDLISPAVVSKTNG 420
DB 361 rTeYvGPtEMiQILAIrAOvEIdMEESlAVGEIGQoTSURhAIdLISpAVvSKtNg 420
QY 421 REKICKADLEEVSGLYDAKSSARLLQEOQERYIT 455
DB 421 reKiCKADleEVsGLyDAKSSARllQEOQeRYIT 455

RESULT 2
AAV72562

ID AAV72562 standard; Protein: 455 AA.

XX AAV72562;

XX 02-MAY-2001 (first entry)

XX Maize RuVB orthologue protein #1.

KW Maize; RuVB orthologue; branch migration; heteroduplex extension;
KW homologous recombination; transformation; transgenic plant.

XX Zea mays.

XX Key Location/Qualifiers

FT Region 40..86 /note="High homology region"

FT Region 214..294 /note="Putative heptad repeat region"

FT Region 297..439 /note="High homology region"

FT Binding-site 297..305 /label="Walker Box_B
/note="ATP binding motif"

XX W0200105975-A1.

XX 25-JAN-2001.

XX 13-JUN-2000; 2000WO-US16271.

XX 16-JUL-1999; 99US-0144112.

XX (PION-) PIONEER HI-BRED INT INC.

XX Mahajan PB;

XX WPI: 2001-159537/16.

XX N-PSDB: AAD02566.

PT Novel maize RuVB nucleic acid useful for modulating levels of maize
PT RuVB in plants, as probes or amplification primers in the detection,
PT quantitation or isolation of gene transcripts -

PS Claim 12: Page 62-63; 87pp: English.

XX The present sequence is a Zea mays RuVB orthologue protein #1. RuVB
CC along with RuVA catalyses the branch migration process, also known as
CC heteroduplex extension, in homologous recombination. RuVB is used for
CC the control of homologous recombination or transformation efficiency in
CC transgenic plants. The RuVB nucleotide may be used as probes or
CC amplification primers for detecting, quantifying or isolating gene
CC transcripts, in detecting deficiencies in the mRNA level during screening
CC for desired transgenic plants, for detecting gene mutations, for
CC monitoring upregulation of expression or changes in enzyme activity, for
CC detecting any number of allelic variants, orthologues or paralogues of
CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
CC also be used for recombinant expression of its encoded polypeptide, or
CC for use as immunogen in preparing and/or screening of antibodies, and in
CC sense or antisense suppression of one or more genes in a host cell,
CC tissue or plant. The RuVB proteins may be used in assays to agonise or
CC antagonise the enzyme function, or as immunogens or antigens for
CC screening antibodies.

XX Sequence 455 AA:

Query Match 99.6%; Score 2253; DB 22; Length 455;

Best Local Similarity 99.3%; Pred. No. 2.1e-189;

Matches 452; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRTEVOSTSKKORATHTTHIKGLIDANGMAIALAGVGQAAREAGLAVDMIRQK 60
DB 1 mlreevgstskkgrlathchikgldangmalaalagvgqaareaglavdmirqk 60

```

QY 61 MAGRAVLGAPPATGKTALALGIAOELGSKVPCPMVGSEVSEVKKTEVLMENFRRAI 120
DB 61 magraavllvppatgktalalgiagelgskvpfcpmvgsevsyevkktevlmenfral 120
QY 121 GLRIKENKEVEGEVTELSPEAEESTTGGYAKSISHVILSLKTVKGTROKLDSSTYDAL 180
DB 121 glrikenkeveyegevtelspeaeesttgyaksishvilskltkvgtkqlkldssiydal 180
QY 181 IKERVAAGDVIIYIANSAGAVKRVGRCDSPATEVDLEAEVYPIPKGEVHKKEIVQDVTL 240
DB 181 ikevavagdviiyiansagavkrvgrcdspatevdlleaevyvipkgevhkkelvqdvltl 240
QY 241 HDLDAANAOPGOGDITLSLMGOMMKPKRTETETKLRQELINKVVRXYDEGIAELVPCVLF 300
DB 241 hlddaanaopgogdditlsimgommkprktekteklrqelinkvvrxydegiaelvpcvlf 300
QY 301 IDEVHMLDIECFSYLRALSPISPIVILATNRGICNVRGTDMTSPHGIPVDLDRLVII 360
DB 301 idevhlmdiecfstylralesplspivilatnrgicnvrtdmtsphgipvdlldrlvii 360
QY 361 RTETVPTPMIOLIAIRAQVEIDMDESLAYIGEIGQTSLRHAIOQLISPAVSYSKTING 420
DB 361 rteetvptpmioliairaqveidmdeslayigeigqtslrhaiqlispasvyskting 420
QY 421 REKICKADLEEVSGLYLDAKSSARLLQEOOERYIT 455
DB 421 rekickadleevsglyldakssarllqegqeryit 455

RESULT 3
AAV72564 standard; Protein; 455 AA.
ID AAV72564
AC AAV72564;
XX
DT 02-MAY-2001 (first entry)
DE Maize RuVB orthologue protein #3.
XX
KW Maize; RuVB orthologue; branch migration; heteroduplex extension;
KW homologous recombination; transformation; transgenic plant.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT Region 40..86
FT Region /note="High homology region"
FT Region 214..294
FT Region /note="Putative heptad repeat region"
FT Region 297..439
FT Region /note="High homology region"
FT Binding-site 297..305
FT Binding-site /label="Walker_Box_B"
FT Binding-site /note="ATP binding motif"
XX
PN W0200105975-A1.
XX
PD 25-JAN-2001.
XX
PF 13-JUN-2000; 2000WO-US16271.
XX
PR 16-JUL-1999; 99US-0144112.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Mahajan PB;
XX
PI WPI: 2001-159537/16.
XX
DR N-PSDB; AAD02368.
XX
PT Novel maize Ruvb nucleic acid useful for modulating levels of maize
PT RuVB in plants, as probes or amplification primers in the detection,

```

```

PT quantitation or isolation of gene transcripts -
XX
PS Claim 12; Page 62-63; 87pp; English.
XX
CC The present sequence is a Zea mays RuVB orthologue protein #3. RuVB
CC along with RuVA catalyses the branch migration process, also known as
CC heteroduplex extension, in homologous recombination. RuVB is used for
CC the control of homologous recombination or transformation efficiency in
CC transgenic plants. The RuVB nucleotide may be used as probes or
CC amplification primers for detecting, quantifying or isolating gene
CC transcripts, in detecting deficiencies in the mRNA level during screening
CC for desired transgenic plants, for detecting gene mutations, for
CC monitoring upregulation of expression or changes in enzyme activity, for
CC detecting any number of allelic variants, orthologues or paralogues of
CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
CC also be used for recombinant expression of its encoded polypeptide, or
CC for use as immunogen in preparing and/or screening of antibodies, and in
CC sense or antisense suppression of one or more genes in a host cell,
CC tissue or plant. The RuVB proteins may be used in assays to agonise or
CC antagonise the enzyme function, or as immunogens or antigens for
CC screening antibodies.
XX
SQ Sequence 455 AA:
XX
Query Match 98.2%; Score 2222; DB 22; Length 455;
Best Local Similarity 97.6%; Pred. No. 1..le-186;
Matches 444; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
QY 1 MRLEEVOSTSKKRIATHYTHIKGLGDANGMAIALAGFVGQAARAAAGLAVDMIROKK 60
DB 1 mrleevostskkriathythikglgdangmampiaagfvgqaaraaagldvdmirgk 60
QY 61 MAGRAVLGAPPATGKTALALGIAOELGSKVPCPMVGSEVSEVKKTEVLMENFRRAI 120
DB 61 magraavllvppatgktalalgiagelgskvpfcpmvgsevsyevkktevlmenfral 120
QY 121 GLRIKENKEVEGEVTELSPEAEESTTGGYAKSISHVILSLKTVKGTROKLDSSTYDAL 180
DB 121 glrikenkeveyegevtelspeaeesttgyaksishvilskltkvgtkqlkldssiydal 180
QY 181 IKERVAAGDVIIYIANSAGAVKRVGRCDSPATEVDLEAEVYPIPKGEVHKKEIVQDVTL 240
DB 181 ikevavagdviiyiansagavkrvgrcdspatevdlleaevyvipkgevhkkelvqdvltl 240
QY 241 HDLDAANAOPGOGDITLSLMGOMMKPKRTETETKLRQELINKVVRXYDEGIAELVPCVLF 300
DB 241 hlddaanaopgogdditlsimgommkprktekteklrqelinkvvrxydegiaelvpcvlf 300
QY 301 IDEVHMLDIECFSYLRALSPISPIVILATNRGICNVRGTDMTSPHGIPVDLDRLVII 360
DB 301 idevhlmdiecfstylralesplspivilatnrgicnvrtdmtsphgipvdlldrlvii 360
QY 361 RTETVPTPMIOLIAIRAQVEIDMDESLAYIGEIGQTSLRHAIOQLISPAVSYSKTING 420
DB 361 rteetvptpmioliairaqveidmdeslayigeigqtslrhaiqlispasvyskting 420
QY 421 REKICKADLEEVSGLYLDAKSSARLLQEOOERYIT 455
DB 421 rekickadleevsglyldakssarllqegqeryit 455

RESULT 4
AAV72565 standard; Protein; 456 AA.
ID AAV72565
AC AAV72565;
XX
DT 02-MAY-2001 (first entry)
DE Maize RuVB orthologue protein #4.
XX
KW Maize; RuVB orthologue; branch migration; heteroduplex extension;

```

homologous recombination; transformation; transgenic plant.

Zea mays.

Key Location/Qualifiers

Region 41..87

Region /note= "High homology region"

Region 215..295

Region /note= "Putative heptad repeat region"

Region 298..440

Region /note= "High homology region"

Binding-site 298..306

/label= Walker_Box_B

/note= "ATP binding motif"

MO200105975-A1.

25-JAN-2001.

13-JUN-2000; 2000MO-US16271.

16-JUL-1999; 99US-0144112.

(PION-) PIONEER HI-BRED INT INC.

Mahajan PB;

MP1: 2001-159537/16.

N-PSDB; AAD02569.

Novel maize Ruvb nucleic acid useful for modulating levels of maize Ruvb in plants, as probes or amplification primers in the detection, quantitation or isolation of gene transcripts -

Claim 12: Page 62-63; 87pp; English.

The present sequence is a Zea mays Ruvb orthologue protein #4. Ruvb along with Ruva catalyses the branch migration process, also known as heteroduplex extension, in homologous recombination. Ruvb is used for the control of homologous recombination or transformation efficiency in transgenic plants. The Ruvb nucleotide may be used as probes or amplification primers for detecting, quantifying or isolating gene transcripts, in detecting deficiencies in the mRNA level during screening for desired transgenic plants, for detecting gene mutations, for monitoring upregulation of expression or changes in enzyme activity, for detecting any number of allelic variants, orthologues or paralogues of the gene, or for site directed mutagenesis in eukaryotic cells. It may also be used for recombinant expression of its encoded polypeptide, or for use as immunogen in preparing and/or screening of antibodies, and in sense or antisense suppression of one or more genes in a host cell, tissue or plant. The Ruvb proteins may be used in assays to agonise or antagonise the enzyme function, or as immunogens or antigens for screening antibodies.

Sequence 456 AA;

Query Match 97.7%; Score 2211.5; DB 22; Length 456;

Best Local Similarity 97.4%; Pred. No. 9.2e-186;

Matches 444; Conservative 7; Mismatches 4; Indels 1; Gaps 1;

1 MREEVOSTSKKORITFHRIKGLD-ANGMAIALAGVGGAAREAGLAVDMIRK 59

1 mreevostskkqritathhikylgdngmmpiaafvypgaareagaavdmirqk 60

60 KMAGRAVLLAGPATGKTALALGIAOELGSKVPFCPMVGSEVYSSBKTEVLMENFRRA 119

61 kmaagraallagpatgktaalalgiagelgskvpfcpmvgysevyssbktevlmenfrfa 120

120 IGLIRKENEVYEGEYTESPEABESTTGAKSISHVLSIKTYKQKQLKLDSSITDA 179

121 iglirkenevyegeytespeaesttgysakishvlsiktykqkqlkldpsiyda 180

180 LIKEKAVAGDIYIIEANSKAVKRGCDSPATEYDLAEAEYVP1PKGEVHKKEIVODVT 239

181 llikekavagdiyleanskavkrgcdsfateydlaeaeeyvp1pkgevhkkekivgdt 240

240 LHDLDANAQPOGGODIISLMGOMAKPKRTETETKLRQELNKVNRRIIDEGIAELVGV 299

241 lhdldaanaqpoggodislmgomakprkrtetetklrqelnkvnrriidegiaelv 300

300 FLDEVMHMDICFSLNNALESPLSPYIILATNNGICNVGCTDTPHGIPLVLDLRLVI 359

301 fldevmhdicfslnnalesplspylilatnngicnvgtctdtpghiplvldlrlvi 360

360 IRTETVGTPEMIQILAIRAQUEEIDMDESLAVLGEIGQOTSLRHAQILISPASVSKTN 419

361 irtetvgtpemiqilairaqueeideeslavlgigqotslrhaqilispasvsktn 420

420 GREKICRADLEEVSGLYIDAKSSARLLQEOQERIT 455

421 grekicradleevsglyidakssarllqeqeryit 456

RESULT 5

AAV72566

ID AAV72566 standard; Protein: 438 AA.

AAV72566;

02-MAY-2001 (first entry)

Maize Ruvb orthologue protein #5.

Maize: Ruvb orthologue; branch migration; heteroduplex extension; homologous recombination; transformation; transgenic plant.

Zea mays.

Key Location/Qualifiers

Region 40..69

Region /note= "High homology region"

Region 197..277

Region /note= "Putative heptad repeat region"

Region 280..422

Region /note= "High homology region"

Blinding-site 280..288

/label= Walker_Box_B

/note= "ATP binding motif"

MO200105975-A1.

25-JAN-2001.

16-JUL-1999; 99US-0144112.

13-JUN-2000; 2000MO-US16271.

MP1: 2001-159537/16.

N-PSDB; AAD02570.

Novel maize Ruvb nucleic acid useful for modulating levels of maize Ruvb in plants, as probes or amplification primers in the detection, quantitation or isolation of gene transcripts -

Claim 12: Page 62-63; 87pp; English.

The present sequence is a Zea mays Ruvb orthologue protein #5. Ruvb along with Ruva catalyses the branch migration process, also known as heteroduplex extension, in homologous recombination. Ruvb is used for the control of homologous recombination or transformation efficiency in transgenic plants. The Ruvb nucleotide may be used as probes or

CC amplification primers for detecting, quantifying or isolating gene
CC transcripts, in detecting deficiencies in the mRNA level during screening
CC for desired transgenic plants, for detecting gene mutations, for
CC monitoring upregulation of expression or changes in enzyme activity, for
CC detecting any number of allelic variants, orthologues or paralogues of
CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
CC also be used for recombinant expression of its encoded polypeptide, or
CC for use as immunogen in preparing and/or screening of antibodies, and in
CC sense or antisense suppression of one or more genes in a host cell,
CC tissue or plant. The Ruvb proteins may be used in assays to agonise or
CC antagonise the enzyme function, or as immunogens or antigens for
CC screening antibodies.

Sequence 438 AA:

Query Match 95.3%; Score 2157.5; DB 22; Length 438;
Best Local Similarity 96.3%; Pred. No. 4.8e-181;
Matches 438; Conservative 0; Mismatches 0; Indels 17; Gaps 2;

QY 1 MREEVOSTSKORATHTHTIKGLDANGMATAALAGFVGOAAREAGLAVDMTRQK 60
DB 1 mreevgstskqrhtchthkglglangmalalaagfvg-----gkk 44
QY 61 MAGRAVLLAGPPATGKTALALGIAQELGSKVPPCPMGSEVYSSEVKKTEVIMENFRRAI 120
DB 45 magravllagppatgkatala-glaqelgskvplcpmgsevysvkktevimenfral 103
QY 121 GRIKENKEVEGETELSPPEASTTGGYAKSISHVILKTVKGTQLDLSIYDAL 180
DB 104 grikenekevegetelspeeasttggysakshvilstkvtgkqklidsiyl 163
QY 181 IKEKAVADVITYEANSQAVKRVGRCDSFATERYDLEAEVYPIPKGEVHKKEIVODVTL 240
DB 164 ikekaavadvityeansqavkrvgrcdsfaterydleaevyppikgevhkkelivodvtl 223
QY 241 HHLDAANQPOGGDITLSMGOMKPKRTEITTEKROINKVVRVYDEGIAELVPCVLF 300
DB 224 hhlदानpqggdiltismgommkprkteittekroinkvvrvydegiaelvpcvlf 283
QY 301 IDEVHMLDIECFSYLNRALESPLSPVILATNRGICNVKGTDMSPRGIFVLDLRLVIT 360
DB 284 idevhlmdiecfsynralesplspvilatnrgicnvrgtdmstprgipvldlrlvilt 343
QY 361 RRETYGPTMIQILAIRAQVEIDMEDESLAYLGEIGQOTSRLRAIQLISPAVSKTNG 420
DB 344 rretygptemqilairaveeidmeeslaylgeigqotsrlraiqilispavsktng 403
QY 421 REKICKADLEEVSGLYIDAKSARLLOEOERYTT 455
DB 404 rekickadleevsgilyidaksarllloeqerytt 438

RESULT 6

AAG20673 standard; Protein; 458 AA.

AAG20673;

17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 22955.

KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

XX

PF 25-FEB-2000: 2000EP-0301439.
XX 25-FEB-1999: 99US-0121825.
PR 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 25-MAR-1999: 99US-0126264.
PR 26-MAR-1999: 99US-0126785.
PR 01-APR-1999: 99US-0127462.
PR 06-APR-1999: 99US-0128234.
PR 08-APR-1999: 99US-0128714.
PR 16-APR-1999: 99US-0129845.
PR 19-APR-1999: 99US-0130077.
PR 21-APR-1999: 99US-0130449.
PR 23-APR-1999: 99US-0130510.
PR 23-APR-1999: 99US-0130891.
PR 28-APR-1999: 99US-0131449.
PR 30-APR-1999: 99US-0132048.
PR 30-APR-1999: 99US-0132407.
PR 04-MAY-1999: 99US-0132484.
PR 05-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132486.
PR 06-MAY-1999: 99US-0132487.
PR 07-MAY-1999: 99US-0132683.
PR 11-MAY-1999: 99US-0134256.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0135629.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136382.
PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137502.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139492.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141827.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.

```

PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145193.
PR 22-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145226.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146387.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 05-AUG-1999; 99US-0147360.
PR 06-AUG-1999; 99US-0147363.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

```

```

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 88.1%; Score 1994; DB 21; Length 458;
 Best Local Similarity 85.1%; Pred. No. 1; le-166;
 Matches 387; Conservative 39; Mismatches 29; Indels 0; Gaps 0;

```

QY 1 MRKEVOSTSKORATTHHIGLIDANGAMALACGFVGAAAREAGLAVDMIRKOK 60
DB :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 MAGRAVLGAPATGKTALALGIAOELGSKVPCMWSEVSSPVKTEVLMENFRRAI 120
DB ||:::||||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 GLRIKENKVEYGEVTELSPEEAESTTGGYAKSISHVITSIKTVKGTQLKLDSSYDAL 180
DB ||||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 124 GLRIKETKEVGEVTELSPEETESLTGGYKSIASHVITLKTIVGTHLIDPLIYDAL 183
QY 181 IREKVAAGDVITYEANSAGVRRVGRCDSEATXEDLAEEYPIRGEVHKKEIVDYTL 240
DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 184 IREKVAAGDVITYEANSAGVRRVGRCDSEATXEDLAEEYPIRGEVHKKEIVDYTL 243
QY 241 HDLDAANAPOGGODILSLMGOMKPRKTEITEKLRORINKVNRVIDEGIAELVPGVLF 300
DB qlldaanaarpqgqdlslmgomkprkteitdklrgeinkvnrvidegyaaelvpvylf 303
QY 301 IDEVHMLDIECFSTYNRLLESPFIYVLATNRGICNVRGTDMTSPHGIPVLLDLRLVIT 360
DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 304 IDEVHMLDIECFSTYNRLLESPFIYVLATNRGICNVRGTDMTSPHGIPVLLDLRLVIT 363
QY 361 RTETYGPTEMIOIILAIRAOVEIDMDEESLAVIGEIOQOSTLRRAIOLISASVSKTNG 420
DB ||:| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 364 rtqlydpsemqilaitraqveeltvdeecvlvlgelqtrslshavqlispasivakmg 423
QY 421 REKICKADLEEVSGIYLDKASSARLQBOOEYRTT 455
DB |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 424 rdnlckadievtsllydaksaklllneqekyis 458
DB

```

RESULT 7
AAG20674
ID AAG20674 standard; Protein; 401 AA.
XX
AC AAG20674;
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22956.
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

```
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160860.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 79.0%; Score 1787; DB 21; Length 401;

Best Local Similarity 86.3%; Pred. No. 1.4e-148; Mismatches 22; Indels 0; Gaps 0;

```
Matches 346; Conservative 33; Mismatches 22; Indels 0; Gaps 0;
QY 55 MIROKKNAGRAVLLGPPATGKTALALGIAOELGSKVPCFPMVGSSEVYSSEVKKTEVIME 114
DB 1 mikyxxkmgakalllaagpgygtcatalalgtlsgelgskvpcfgmvgsevsyevkktevlme 60
QY 115 NFRRAIGLRIRKEKKEVEGEVTELSPEASSTTGGYAKSISHVITISLKYVKGTKQLKLDL 174
DB 61 nfttraiglrirkekevegevtelsspeetstlgygysishvltlklvkgtkhklldp 120
QY 175 SIYDALIKERVAAGDVYIIFANGSAVKRVGCDSPFATEVDLEEEVVPPIPKGVHKKKEEL 234
DB 121 tlydallikekvavgdvlylfeangsaavkryvgsdafateideeevvpipkgevnkkkel 180
```

```
QY 235 VODVTLHDLDAANAPOGGODIILMCGOMKPRKTEITTEKLRQENKVNRYIDEGLAEL 294
DB 191 vgdvtrlglddaanaipgggddilsmgqmkprkteittdkrlrqenkvnrlyidegvael 240
QY 295 VPGVLEIDEVHMDIECFSYLNRALESPLSPVTLATNRGICNVGTDMSPHGIPVDL 354
DB 241 vpgvlfidevhlmdiecfyslnralesslpvltfatnrgvncvrgtdmpspgypdl 300
QY 355 DRVYIRRTETGPTPEMIOILAIRAQVEIIMDEESLAVLGEIGQSTSLRAIQLISPAV 414
DB 301 drvylirtqlydpsemqqlaliraveeltvdeclvlgelqtrtslrnavqlispasl 360
QY 415 VSKTNGREKICKADLEEVSGLYIDAKSSARLLPOOERYYT 455
DB 361 vakmgtrdnickadieevtslyidakssakllneqgekyls 401

RESULT 8
AAG20675
ID AAG20675 standard; Protein; 395 AA.
XX
AC AAG20675;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22957.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
```


PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157173.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.

Query Match 77.8%; Score 1761; DB 21; Length 395;
Best Local Similarity 86.3%; Pred. No. 2,6e-146;
Matches 341; Conservative 32; Mismatches 22; Indels 0; Gaps 0;

```

QY 61 MAGRAVLLAGPPATGKTLALGIAOELGSKVPCPMVGSEVYSSPVKTEVLMENFRRAI 120
   1 magkalllagppgltalalaglsqelskvpfcpmvgsevyssvkvktemenfrrai 60
DB 121 GLRIKENKEVEYEGEVELTSPPEAEESTTGGYAKSISHVIISLKTGKTKOLKLDSSITYDAL 180
   61 glriketkeveyegeveltsppestestlsgyaksishviltktvgkthklkldpiyal 120
QY 181 IREKAVAGDVITYEANGAVKRVGCDSEFATRYDLEAEYVPIPKGEVHKKEIYODVTL 240
   121 irekavagdvityeangavkrvgrcdsefatrydleaeyvpipkgevhkkekelyodvcl 180
DB 241 HDLDAANAOPGOGDIISLMGOMMKPRKTEITKLEKROEINKVVRNRYIDEGIAELVGVLF 300
   181 qdlldanaarpggqgdllsmgmmkprkteleitkrlrgeinkvvrnyidegyaelvpgvlf 240
QY 301 IDEVHMLDIECFSYLNRALESPLSPVILATNRGICNVGTDMTSPHGIPVDLDRVLI 360
   241 idevhlmdiecfstylralesplspviltatnrgicnvgtdmtsphgipvdldrlvli 300
QY 361 IRTFTYGTPTMIOITLIRAOVEEIDMDESLAYLGIGOOTSLRHAIOILSPASVSKTNG 420
   301 irfttygtptmioitliraoveeidmdeslaylgigootslrhaioilspasvsktng 360
DB 421 REKICKADLEEVSGLYLDKSSARLLOEOERYI 455
   361 rdnickadieevtslyldkssakllhegqekyis 395

```

RESULT 9

AAW74417
ID AAW74417 standard; Protein; 456 AA.

```

XX AC AAW74417;
XX DT 10-MAY-1999 (first entry)
XX DE TIP49 protein sequence #2.
XX KM TIP49; TATA binding protein; TBP binding protein;
XX KM DNA transcription control.
XX OS Rattus sp.
XX PN WO9900419-A1.
XX PD 07-JAN-1999.
XX PF 25-JUN-1998; 98WO-JP02836.
XX PR 27-JUN-1997; 97JP-0187398.
XX PA (SUME ) SUMITOMO ELECTRIC IND CO.
XX PI Kishimoto T, Makino Y, Niwa S, Tamura T;
XX DR WPI; 1999-095682/08.
XX DR N-PSDB; AAX18198.
XX PT TBP-binding protein with DNA helicase and ATPase activities - gene
XX PT encoding it, and antibodies recognising it.
XX PS Claim 2; Page 41-43; 64pp; Japanese.

```

CC This sequence is the TATA-binding protein (TBP) binding protein, designated TIP49, of the invention. TIP49 and its associated nucleic acids and antibodies are useful in investigation of the process of DNA

CC transcription control by TBP in vivo, and in detection of the blocking of
CC DNA transcription.

SQ Sequence. 456 AA;

Query Match 76.2%; Score 1723.5; DB 20; Length 456;
Best Local Similarity 72.5%; Pred. No. 6,3e-143;
Matches 330; Conservative 73; Mismatches 51; Indels 1; Gaps 1;

```

QY 1 MREIEVOSTSKKORLATHTHIKGLGDANGMALALAGFYGQAANAAGLAVDMIROKK 60
   1 mkleevostskkqltashshvkglgidesglakqasglvgenareacgvivellksk 60
DB 61 MAGRAVLLAGPPATGKTLALGIAOELGSKVPCPMVGSEVYSSPVKTEVLMENFRRAI 120
   61 magraavltagppgltalalaglsqelskvpfcpmvgsevyssvkvktemenfrrai 120
QY 121 GLRIKENKEVEYEGEVELTSPPEAEESTTGGYAKSISHVIISLKTGKTKOLKLDSSITYDAL 180
   121 glriketkeveyegeveltsppestestlsgyaksishviltktvgkthklkldpsifeel 180
DB 181 IREKAVAGDVITYEANGAVKRVGCDSEFATRYDLEAEYVPIPKGEVHKKEIYODVTL 240
   181 qkerveagdvityeangavkrvgrcdsefatrydleaeyvpipkgevdvnhkkelvgdvl 240
QY 241 HDLDAANAOPGOGDIISLMGOMMKPRKTEITKLEKROEINKVVRNRYIDEGIAELVGVLF 300
   241 hdlvdaanaopgogdiislmgmmkprkteleitkrlrgeinkvvrnyidegyaelvpgvlf 300
DB 301 IDEVHMLDIECFSYLNRALESPLSPVILATNRGICNVGTDMTSPHGIPVDLDRVLI 360
   301 idevhlmdiecfstylralesplspviltatnrgicnvgtdmtsphgipvdldrlvli 360
QY 361 IRTFTYGTPTMIOITLIRAOVEEIDMDESLAYLGIGOOTSLRHAIOILSPASVSKTNG 420
   361 irfttygtptmioitliraoveeidmdeslaylgigootslrhaioilspasvsktng 420
DB 421 REKICKADLEEVSGLYLDKSSARLLOEOERYI 454
   421 gkdsiekehveiseltyldkssaklladqdkym 455

```

RESULT 10

AAW74416
ID AAW74416 standard; Protein; 456 AA.

```

XX AC AAW74416;
XX DT 10-MAY-1999 (first entry)
XX DE TIP49 protein sequence.
XX KM TIP49; TATA binding protein; TBP binding protein;
XX KM DNA transcription control.
XX OS Rattus sp.
XX PN WO9900419-A1.
XX PD 07-JAN-1999.
XX PF 25-JUN-1998; 98WO-JP02836.
XX PR 27-JUN-1997; 97JP-0187398.
XX PA (SUME ) SUMITOMO ELECTRIC IND CO.
XX PI Kishimoto T, Makino Y, Niwa S, Tamura T;
XX DR WPI; 1999-095682/08.
XX DR N-PSDB; AAX18198.

```

PT TBP-binding protein with DNA helicase and ATPase activities - gene
PT encoding it, and antibodies recognising it.

PS Claim 1, Page 37-39; 64pp: Japanese.

XX This sequence is the TATA-binding protein (TBP) binding protein,
CC designated TBP4, of the invention. TBP4 and its associated nucleic
CC acids and antibodies are useful in investigation of the process of DNA
CC transcription control by TBP in vivo, and in detection of the blocking of
CC DNA transcription.

SO Sequence 456 AA;

Query Match 76.1%; Score 1722.5; DB 20; Length 456;
Best Local Similarity 72.3%; Pred. No. 7.8e-143;
Matches 329; Conservative 74; Mismatches 51; Indels 1; Gaps 1;

OY 1 MRIEVOSTSKORATHTHKGIGDANGMALAAGFVQAAREAGLAVDMIRKK 60
DB 1 mkleevkstktqriashshvkqigldesglakgaasglvqgenareacgivelleskk 60
OY 61 MAGRAVLLAGPATGKTALALGIAOELGSKVPCPMVGSSEYSEVKKTEVLMENFRRAI 120
DB 61 magravllagppgqgkatalalagelgskvpcpmvgsseystelktevlmennfral 120
OY 121 GLRIKENKEVEGEVTELSPEAEESTGTGVAKSISHVIISLKTIVKGTOKLDSSTYDAL 180
DB 121 glriketkevegevteltpcelempmgysktishvliigltakgtqklidpsifesi 180
OY 181 IKERKAVGDVLYTEANSAGVRGRCDSFATEYDLAEAEVYPIPKGEVHKKEIVODVTL 240
DB 181 qkerveagdvlyteansgavrkgrcdyatefdlaeevypipgdkhkkeliqdvtl 240
OY 241 HDLDANAPOGODILSLMGOMMKPRKTEITTEKLRKOEINKVNRKYIDEGIAELVPGVLF 300
DB 241 hldldanarpogqgdlismmgqmkpkteitdtklrgelinkvnyldgvaeeilpgvlf 300
OY 301 IDEVIMLDIECFSTYNRLAESPLSPYIVLANNRGICNVKGT-DMTSPHGIPTVDLRLVI 359
DB 301 idevimldiecfstynrlaesplspyivlannrgicnvgktdmtspghiptvdlrlvml 360
OY 361 IRTETYPTEMIQILAIRAOVEIDMEESLAYIGEIGQOTSLRHAQILISPAVSKTN 419
DB 361 irtetypkemqkikiraqegelniseeanhngelstktltlrysvqlipnallakin 420
OY 420 GREKICKADLEEVSGLYLDANSSARLLQEOQERYI 454
DB 421 gkdsiekehveiselldaksakilladgdkym 455

RESULT 11

ID AAW93945 standard; Protein; 456 AA.

AAW93945;

30-JUN-1999 (first entry)

Human regulatory molecule HRM-1 protein.

KW Human regulatory molecule: HRM-1; cytosolic activity; immune modulator;
KW transcription factor; enhancer; cell proliferation stimulation; cancer;
KW treatment; microarray; detection; diagnosis; cell proliferation disease;
KW leukemia; lymphoma; myeloma; adenocarcinoma; sarcoma; bladder; bone;
KW brain; lung; liver; ovary; skin; teratocarcinoma; immune response;
KW allergy; asthma; diabetes; multiple sclerosis; Grave's disease;
KW myasthenia gravis.

OS Homo sapiens.

WO9915658-A2.

XX

PD 01-APR-1999.

XX 22-SEP-1996; 98WO-US19839.

XX 23-SEP-1997; 97US-0933750.

XX (INCY-) INCYTE PHARM INC.

PI Au-Young J, Bandman O, Guegler KJ, Hillman JL, Lal P;

PI Shah P, Yue H;

DR WPI; 1999-254710/21.

DR N-PSDB; AAX24059.

XX New human regulatory molecules

XX Claim 1, Page 60-61; 76pp: English.

CC This invention describes novel human regulatory molecules (HRM) which
CC have cytosolic activity and act as immune modulators, transcription
CC factors or enhancers. The HRMs can be used to stimulate cell
CC proliferation. Antagonists and agonists of the proteins of the invention
CC can be used to treat cancer. The encoding nucleic acids can be used in
CC microarrays to detect polynucleotides (and their expression levels) that
CC encode HRMs in a biological sample. The HRMs and microarrays can be used
CC to diagnose, treat or prevent cell proliferation diseases especially cancer,
CC e.g., leukemia, lymphoma, myeloma, adenocarcinoma, sarcoma, cancer of e.g.,
CC bladder, bone, brain, lung, liver, ovary, skin, etc., teratocarcinoma, or
CC to treat or prevent immune responses e.g., allergies, asthma, diabetes,
CC multiple sclerosis, Grave's disease or myasthenia gravis.

SO Sequence 456 AA;

Query Match 76.1%; Score 1721.5; DB 20; Length 456;
Best Local Similarity 72.5%; Pred. No. 9.5e-143;
Matches 330; Conservative 72; Mismatches 52; Indels 1; Gaps 1;

OY 1 MRIEVOSTSKORATHTHKGIGDANGMALAAGFVQAAREAGLAVDMIRKK 60
DB 1 mkleevkstktqriashshvkqigldesglakgaasglvqgenareacgivelleskk 60
OY 61 MAGRAVLLAGPATGKTALALGIAOELGSKVPCPMVGSSEYSEVKKTEVLMENFRRAI 120
DB 61 magravllagppgqgkatalalagelgskvpcpmvgsseystelktevlmennfral 120
OY 121 GLRIKENKEVEGEVTELSPEAEESTGTGVAKSISHVIISLKTIVKGTOKLDSSTYDAL 180
DB 121 glriketkevegevteltpcelempmgysktishvliigltakgtqklidpsifesi 180
OY 181 IKERKAVGDVLYTEANSAGVRGRCDSFATEYDLAEAEVYPIPKGEVHKKEIVODVTL 240
DB 181 qkerveagdvlyteansgavrkgrcdyatefdlaeevypipgdkhkkeliqdvtl 240
OY 241 HDLDANAPOGODILSLMGOMMKPRKTEITTEKLRKOEINKVNRKYIDEGIAELVPGVLF 300
DB 241 hldldanarpogqgdlismmgqmkpkteitdtklrgelinkvnyldgvaeeilpgvlf 300
OY 301 IDEVIMLDIECFSTYNRLAESPLSPYIVLANNRGICNVKGT-DMTSPHGIPTVDLRLVI 359
DB 301 idevimldiecfstynrlaesplspyivlannrgicnvgktdmtspghiptvdlrlvml 360
OY 361 IRTETYPTEMIQILAIRAOVEIDMEESLAYIGEIGQOTSLRHAQILISPAVSKTN 419
DB 361 irtetypkemqkikiraqegelniseeanhngelstktltlrysvqlipnallakin 420
OY 420 GREKICKADLEEVSGLYLDANSSARLLQEOQERYI 454
DB 421 gkdsiekehveiselldaksakilladgdkym 455

RESULT 12

AAV73986

```
ID AAY73986 standard; Protein; 418 AA.
XX
XX AAY73986;
XX
XX
DT 14-MAR-2000 (first entry)
XX
XX Human prostate tumor EST fragment derived protein #173.
DE
XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
XX treatment.
XX
XX Homo sapiens.
OS
XX DE19820190-A1.
XX
XX 04-NOV-1999.
XX
XX 28-APR-1998; 98DE-1020190.
XX
XX 28-APR-1998; 98DE-1020190.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI; 1999-621386/54.
XX
XX N-PSDB; AAZ52915.
XX
XX New human nucleic acid sequences from pancreatic tumors, and related
XX proteins -
XX
XX Claim 23; Page 381; 502pp; German.
XX
XX This invention describes novel polypeptides and their encoding nucleic
XX acids derived from human pancreatic tumor tissue which have cytostatic
XX activity. The sequences are also useful in producing pharmaceutical
XX compositions for treatment of pancreatic tumors. AAY73814-174252
XX represent protein fragments encoded by the human pancreatic tumor cDNA
XX library derived expressed sequence tag (EST) sequences represented in
XX AA52858-253014.
XX
XX
XX Sequence 418 AA;
SQ
Query Match 68.1%; Score 1541.5; DB 20; Length 418;
Best Local Similarity 73.8%; Pred. No. 5.3e-127;
Matches 296; Conservative 61; Mismatches 43; Indels 1; Gaps 1;
QY 47 EAAGLAIVMIRQKMGAGAVLLAPPRATGKATLALGIAQELGSKVPPCPMVGSEVSSSEV 106
DB 12 qacgvivelliskkmegavllaqpgtlyktalalaiaqelgskvppcpmvsevystel 71
QY 107 KTEVLMENFRAIGLRIKENKEVEGEVTELSPEEAESTTGAKSISHVITSLKTVKG 166
DB 72 ktevlmenfrraiglrilkenkevegevtelcpecelemmgvgkctishvllgikakg 131
QY 167 TKQLKLDSSIVDALIKERKAVAGDVITYEANSAGAVKRVGRCDSPATEYDLAEAEVPIPKG 226
DB 132 tkqlklpslfselqkerveagdvityeansgavkrgcdtyatetdlaeaevyvplpg 191
QY 227 EVHKKKEIVDDVTLHDLDANNAOQGGODITLSGMKMKRKTETPKLQELINKVYNRK 286
DB 192 dvhkkkeivddvtlhdldannapqgqdlismgqimkpkctetlcklrgelinkvynky 251
QY 287 IDEGIAELVGVLEIDVENHMDIECFSYLNRALESPIVILATNNGICNVNGT-DMTS 345
DB 252 idegiaelvglvleidevenhmdiecfylnralespiivilasnrgncvirlgledits 311
QY 346 PHGIPVULDLRLVYIIRTEYGPTEMTQIILAIRAOVEIDMDSESLAVLGEIQQOTSILRNA 405
DB 312 phgipvuldrlvliirteygpptemioiilaiaroveidmdeslavlgelqqotsilrns 371
QY * 406 IQLISPAVSASVSKNGREKICKADLEVEVSGLYLDAKSSARLL 446
```

```
Db 372 vqlltpanllakngkdsiekheveiselvfydaksakll 412
RESULT 13
AAB21093
ID AAB21093 standard; Protein; 463 AA.
XX
XX
XX AAB21093;
XX
XX 22-DEC-2000 (first entry)
XX
XX Human HEL50 DNA helicase.
DE
XX
XX Human; HEL50; DNA helicase; TTP49 homologue; EST AA374580;
XX expressed sequence tag; antibody; antisense therapy; diagnosis;
XX helicase-associated disease; genetic disorder; tumour; cancer.
XX
XX Homo sapiens.
OS
XX
XX WO200047731-A1.
XX
XX 17-AUG-2000.
XX
XX 22-NOV-1999; 99WO-JP06519.
XX
XX 10-FEB-1999; 99JP-0033062.
XX
XX (SUME) SUMITOMO ELECTRIC IND CO.
XX
XX Tamura T;
XX
XX WPI; 2000-543587/49.
XX
XX N-PSDB; AAA90316.
XX
XX DNA helicase of human origin, antibodies to it and DNA encoding it for
XX investigation diagnosis and treatment of genetic disorders involving
XX helicase gene expression -
XX
XX Claim 1; Page 39-40; 49pp; Japanese.
XX
XX This sequence represents a novel human DNA helicase, HEL50, which has
XX homology with a part of human TTP49. cDNA encoding human HEL50 was
XX initially identified in an EST (expressed sequence tag) database as
XX CC EST AA374580. The invention relates to human and yeast HEL50 DNA
XX CC helicases (AAB21093, AAB21094), and to cDNAs encoding them (AAA90316,
XX CC AAA90317). It also encompasses HEL50 antisense oligonucleotides and
XX CC anti-HEL50 antibodies. HEL50 nucleotides, proteins and antibodies may be
XX CC used for the investigation, diagnosis and treatment of diseases with
XX CC which DNA helicase is associated, such as genetic disorders involving
XX CC expression of the DNA helicase gene occur.
XX
XX Sequence 463 AA;
SQ
Query Match 43.0%; Score 973; DB 21; Length 463;
Best Local Similarity 44.0%; Pred. No. 4.9e-77;
Matches 200; Conservative 97; Mismatches 144; Indels 14; Gaps 4;
QY 2 RLEEVOSTSKKRIATHTHIKGLDANGMATALAAEFGAAREAGLAVDMIRQKM 61
DB 9 rleevostskkriathtthikglidangmatalaaefvgaareaglavdmirqkm 68
QY 62 AGRAYVLLGPPATGTGATLALGIAQELGSKVPPCPMVGSEVSSSEVKKTEVLMENFRAIG 121
DB 69 agravvllgppatgtgatalalgaqelgskvppcpmvgsesvssvkktevlmenfrraig 128
QY 122 LRIKENKEVEGEVTELSPEEAESTTGAVKASISHVITSLKTVGTRKQLKLDSSIVDALI 181
DB 129 vrikenkevegevtelcpeeaesttgavkasisshvitslktvgtkqlkldssivdali 183
QY 182 KEKVAAGDVITYEANSAGAVKRVGRCDSPATEYDLAEAE-VYPIPKGEVHKKKEIVDDVT 239
```

```

Db 184 kdkvgagvdtidkatgyskigrsfrardydamsqtkftvgcpdgelqkrkevhtvs 243
OY 240 LHDLDANAAGOGGDDIISLMGOMKPKRTITEKROELNKVYNRVIDEGIAELVPGVL 299
Db 244 lhelvlnstrtqg---flal-----fsgdtgeiksevregeqnakvaewreegkaelipgvl 296
OY 300 FIDEVHMDICEFSYLNALRESPLPIVILATNRGICNVRGTDMTSPHGIPVDLDRLYI 359
Db 297 fidevhdmdicesfslmralsedmapvlmatnqrltrtqyspsphgipdlldrlfll 356
OY 360 IRTETVGPTEMIQILAIRAQVEEIDMDEESLAYLGEIGOOTSLRHAIQILISPASVSEKTN 419
Db 357 vsttprsekdtkqllrrceeedvemsedaytvltrigletsrlayaqlitaaslvrctr 416
OY 420 GREKICRADLEFVSGLYIDAKSSARLLOEOERYI 454
Db 417 kgtewyvddikrvyslfldearstqymkeyqdafl 451

```

RESULT 14

AAB21094
ID AAB21094 standard; Protein; 471 AA.

```

XX AC AAB21094;
XX DT 22-DEC-2000 (first entry)
XX DE Yeast HEL50 DNA helicase.
XX KW Yeast; HEL50; DNA helicase; TIP49 homologue; antibody; antisense therapy;
XX KM diagnosis; helicase-associated disease; genetic disorder; tumour; cancer.
XX OS Saccharomycetes cerevisiae.
XX WO200047731-A1.
XX PN 17-AUG-2000.
XX PD 22-NOV-1999; 99WO-JP06519.
XX PF 10-FEB-1999; 99JP-0033062.
XX PR (SUME ) SUMITOMO ELECTRIC IND CO.
XX PA Tamura T;
XX PI WPI; 2000-543587/49.
XX DR N-PSDB; AAA90317.
XX XX
XX PT DNA helicase of human origin, antibodies to it and DNA encoding it for
XX PT investigation diagnosis and treatment of genetic disorders involving
XX PT helicase gene expression -
XX PS Claim 4; Page 44-45; 49pp; Japanese.
XX CC This sequence represents a novel yeast DNA helicase, HEL50, which
XX CC has homology with a part of human TIP49. The invention relates to
XX CC human and yeast HEL50 DNA helicases (AAB21093, AAB21094), and to cDNAs
XX CC encoding them (AAA90316, AAA90317). It also encompasses HEL50 antisense
XX CC oligonucleotides and anti-HEL50 antibodies. HEL50 nucleotides, proteins
XX CC and antibodies may be used for the investigation, diagnosis and
XX CC treatment of diseases with which the DNA helicase is associated, such as
XX CC genetic disorders involving variant forms of the gene, and tumours in
XX CC which altered levels of expression of the DNA helicase gene occur.
XX SQ Sequence 471 AA.

```

Query Match 41.2%; Score 932.5; DB 21; Length 471;
Best local similarity 44.6%; Pred. No. 1.8e-73;
Matches 198; Conservative 88; Mismatches 139; Indels 19; Gaps 6;

```

OY 15 IATHTHKGIGLDANGMALAAGFVGQAAREAGLAVDMTROKKMACRAVILLAPPYT 74
Db 20 laahshltglgldentlqprptsegmwyqlqarraagvllkmwyngflaagravlvagppst 79
OY 75 GKTAALGIAEOEGSVKPCPMWSEVSEVSEVSEVSEVSEVSEVSEVSEVSEVSEVSEVSEV 134
Db 80 gktalamvsgslgkvdpfltaigseifslslelstealqatrlsrgikikeetellege 139
OY 135 VTELSPDEASTTGGYAKSISHVILSKTVYGTQKQLDSSIYDALIKERKAVGVYIYE 194
Db 140 vtelidqld--rstlqghkqg---klfiktmetiylgnkmi dgltkktkviagdvlsid 193
OY 195 ANGAARKVRGRCSFATVDLEA-----EEVPIPKGEVKKKKEIVODVTLLHDLAANAOP 250
Db 194 kasgkltiklgr--stasrtdydamgadrltfvqpegeqlkrtlvhtvs lhelvlnstr 251
OY 251 OGQODILSLMGOMKPKRTITEKROELNKVYNRVIDEGIAELVPGVLFIDEVHMDICE 310
Db 252 qg---flal-----ftgdqgelisevrdqnlkvaewkeegkaeivpgvl fidevhdmdice 304
OY 311 CFSYLNRALESPLPIVILATNRGICNVRGTDMTSPHGIPVDLDRLYIIRTEVGPTEEM 370
Db 305 cfsflnraledfaplvmatnrgvsktrgtnyksphgipdlldrstlittksynegei 364
OY 371 IQILAIRAQVEEIDMDEESLAYLGEIGOOTSLRHAIQILISPASVSKTNGREKICKADIE 430
Db 365 klislrageeevelssdailltkrvetslrysnnlsvaqjlamkrkntlvvedvk 424
OY 431 EVSGLYIDAKSSARLLOEOERYI 454
Db 425 rayllflidsarsvkvygenesqyl 448

```

RESULT 15

AAB80611
ID AAB80611 standard; Protein; 334 AA.

```

XX AC AAB80611;
XX DT 02-MAY-2001 (first entry)
XX DE Environmental stress tolerant protein SEQ ID 8.
XX KW Environmental stress resistance; salt; heat; desert; transgenic plant.
XX KM Bruguiiera sezangula.
XX OS
XX OS Bruguiiera sezangula.
XX PN WO200106006-A1.
XX PD 25-JAN-2001.
XX PF 19-JUL-2000; 2000WO-JP04862.
XX PR 19-JUL-1999; 99JP-0235910.
XX PR 24-MAR-2000; 2000JP-0085377.
XX PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX XX Yamada A, Ozeki Y, Saito T;
XX PI WPI; 2001-147355/15.
XX DR N-PSDB; AAF74190.
XX XX
XX PT Screening method to obtain DNA encoding environmental stress resistance
XX PT factor, useful for producing transgenic plants resistant to
XX PT environmental stress -
XX PS Claim 22; Page 86-87; 167pp; Japanese.
XX CC Polynucleotide sequences AAF74187 - AAF74218 encode proteins
XX CC AAB80608 - AAB80639, which impart environmental stress resistance. The
XX CC invention relates to a method for identifying DNA encoding proteins
XX CC imparting environmental stress resistance. The method comprises inserting

```

CC cDNA from a library originating in a salt-resistant organism into a host
CC cell, culturing the transformants under conditions in which the
CC untransformed host does not grow well, and selecting for viable clones.
CC The method is useful for obtaining DNA encoding environmental stress
CC resistance factors. The DNA encoding proteins conferring environmental
CC stress resistance, can be used in the production of plants resistant to
CC environmental stress, which can be cultivated in unfavourable
CC environments such as deserts, salt damaged ground, cold regions and the
CC oceans. They can be used for increasing the area of land covered by green
CC plants, and desert greening and afforestation, in order to counter the
CC effects of the increase in atmospheric carbon dioxide concentration. PCR
CC primers AAF74219 and AAF74220 are used in an example illustrating the
CC method of the invention.

SQ Sequence 334 AA;

Query Match	27.9%	Score 630.5	DB 22	Length 334
Best Local Similarity	43.4%	Pred. No. 3.5e-47		
Matches 141, Conservative	67	Mismatches 104	Indels 13	Gaps 5

```
Oy 132 EGEVTELSPEAEASTTGCGYAKSLSHVITSLKPTVYGTQKQLDSSLIDALIREKVAADV 191
    |||| : : : : : |||| : : : |||| ||||
Db 2 egevevevgidr-pavtgaasktgc--l|l|k|t|e|t|e|v|y|d|g|a|m|i|e|a|l|g|k|e|k|v|g|s|g|v|i 57
```

Qy 192 YEANSQAVKRRGRCDSEFATEYDLEAE--YVPKPGEVAKKKKEIQDVTIHLDDANAQ 249

Db 58 aIdkasqkItlgrfsfsrSdydampgqvfvqcpdgelqkrkevvhcvsIheIdvInsr 117

Qy 250 PGGGDDILSLMGQMMKPRKTEIEKLRQINKVNRIDEGIAELTPGVLFIDEVHMADI 309
 118 tqg---flal----ftgdtgeiraeavreqidtkvaewreegkaeiypvlfidevhmadi 170

QY 310 ECFSYNRALESPLSPVIAIATNGICNRGTDMSPHGIPVDLDRVIAIETEVGPLE 369
 ||||:|||||: :|||:|||||:|||||:| | : |
 Db 171 ecfsflnralemaplvlvatngitttgrtnyksphgipdlldrl|||ttgpyekde 230

Db 231 lrklldlrqeedvemaeeakallthigetslryahllltaaalocqkrkglveterd 290

QY	430	EEVSGHYLDAAKSSARLLDEQDERK1	434
		:: :: :: :	
DB	291	sragn1fldvkrtstqylleygnqym	315

Search completed: November 13, 2001, 07:01:53
Job time: 359 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2001, 09:56:10 ; Search time 100.59 Seconds

(without alignments)
4304.861 Million cell updates/sec

Title: US-09-589-510-3

Perfect score: 1912

Sequence: 1 accacgcgcgcgaattt.....aaaaaaaaaaaaaaaaaa 1912

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PCRTUS_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	612.4	32.0	1750	2 US-08-933-750C-54	Sequence 54, Appl
2	612.4	32.0	1750	3 US-09-234-613-54	Sequence 54, Appl
3	112.6	5.9	703	4 US-08-998-416-1005	Sequence 1005, Ap
4	55.8	2.9	7218	1 US-08-232-463-14	Sequence 14, Appl
5	49.8	2.6	1474	4 US-08-821-994-64	Sequence 64, Appl
6	49.4	2.6	4931	4 US-08-726-320-2	Sequence 2, Appl
7	49.4	2.6	4931	4 US-09-208-716-2	Sequence 7, Appl
8	49.4	2.6	966	1 US-08-514-014-7	Sequence 7, Appl
9	49.4	2.6	966	2 US-08-833-823-7	Sequence 7, Appl
10	48.8	2.6	347	1 US-08-104-072B-2	Sequence 2, Appl
11	48.8	2.6	1117	4 US-09-247-373B-33	Sequence 33, Appl
12	48.8	2.6	3777	3 US-09-121-321-15	Sequence 15, Appl
13	48.8	2.6	3777	4 US-08-933-803A-15	Sequence 15, Appl
14	48.8	2.6	7774	3 US-08-956-307B-11	Sequence 11, Appl
15	48.8	2.6	7774	4 US-08-821-994-63	Sequence 63, Appl
16	47.8	2.5	1441	4 US-08-924-747-25	Sequence 25, Appl
17	47.2	2.5	991	3 US-09-247-373B-25	Sequence 25, Appl
18	47.2	2.5	991	4 US-09-296-715-25	Sequence 25, Appl
19	47.2	2.5	991	4 US-08-592-214A-1	Sequence 1, Appl
20	47.2	2.5	1215	1 US-09-149-976-1	Sequence 1, Appl
21	47.2	2.5	1215	1 US-08-101-593-1	Sequence 1, Appl
22	47.2	2.5	2760	1 US-08-524-757-1	Sequence 1, Appl
23	46.8	2.4	458	1 US-08-361-467B-4	Sequence 4, Appl
24	46.8	2.4	1046	1 US-08-484-332C-4	Sequence 4, Appl
25	46.8	2.4	1046	1 US-08-499-215-1	Sequence 1, Appl
26	46.6	2.4	6379	1 US-08-471-717-1	Sequence 1, Appl
27	46.6	2.4	940	2 US-08-471-717-1	Sequence 1, Appl

28	46.4	2.4	509	4 US-09-030-607-202	Sequence 202, App
29	46.4	2.4	1512	2 US-08-909-965C-8	Sequence 8, Appl
30	46.4	2.4	3527	2 US-08-909-965C-7	Sequence 7, Appl
31	46.4	2.4	1315	4 US-09-721-822A-10	Sequence 10, Appl
32	46.4	2.4	1364	1 US-08-265-087-3	Sequence 3, Appl
33	46.4	2.4	1364	1 US-08-621-493-3	Sequence 3, Appl
34	46.4	2.4	1364	2 US-08-965-688-3	Sequence 3, Appl
35	46.4	2.4	1364	4 US-09-260-173-3	Sequence 3, Appl
36	45.8	2.4	857	3 US-08-460-040-1	Sequence 1, Appl
37	45.8	2.4	882	2 US-08-909-965C-9	Sequence 9, Appl
38	45.8	2.4	1172	1 US-07-945-288-9	Sequence 9, Appl
39	45.8	2.4	1172	1 US-08-462-831-9	Sequence 9, Appl
40	45.8	2.4	1172	1 US-08-461-809-9	Sequence 9, Appl
41	45.8	2.4	1172	1 US-08-461-441-9	Sequence 9, Appl
42	45.8	2.4	1172	5 PCT-US93-08518-9	Sequence 9, Appl
43	45.6	2.4	688	6 5498694-3	Patent No. 5498694
44	45.6	2.4	2209	1 US-08-514-014-1	Sequence 1, Appl
45	45.6	2.4	2209	2 US-08-833-823-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-933-750C-54

; Sequence 54, Application US/08933750C

; Patent No. 5932442

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Bandman, Olga

; APPLICANT: Shah, Purvi

; APPLICANT: Au-Yang, Janice

; APPLICANT: Yue, Henry

; APPLICANT: Guegler, Karl J.

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES

; NUMBER OF SEQUENCES: 98

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/933,750C

FILING DATE: September 23, 1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0356 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

SEQUENCE CHARACTERISTICS:

LENGTH: 1750 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: HMCINOT01

CLONE: 9476
US-08-933-750C-54

Query Match 32.0%; Score 612.4; DB 2; Length 1750;
Best local similarity 66.0%; Pred. No. 7.4e-134;
Matches 902; Conservative 0; Mismatches 461; Indels 3; Gaps 1;

94 atgagatcgaagagtgatcagtcgaagaaagcagcagtcgacccacacccac 153
DB ATGAAGATTGAGAGGTGAGAGAGCAGTACGAGAGCAGCAGCAGCAGCAGC 141
154 atcaagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 213
DB GTGAAGAGGTGAGAGGTGAGAGAGCAGAGCAGCAGCAGCAGCAGCAGC 201
214 ggcagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 273
DB GGCAGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 261
274 atgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 333
DB ATGAGCTGGAAGAGCTGCTTGTGAGCAGCAGCAGCAGCAGCAGCAGC 321
262 ATGAGCTGGAAGAGCTGCTTGTGAGCAGCAGCAGCAGCAGCAGCAGC 321
334 ctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 393
DB CTGAGCTGAGAGAGCTGCTTGTGAGCAGCAGCAGCAGCAGCAGCAGC 381
394 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 453
DB GTTACTGCACTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 441
454 gtttcgataaagagaaagagagagagagagagagagagagagagagag 513
DB GGGCTGGAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 501
442 GGGCTGGAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 501
514 gaagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 573
DB TGTGAG 561
502 TGTGAG 561
574 ttaagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 633
DB CTCAAAAG 621
562 CTCAAAAG 621
634 atcaagagagagcagcagcagcagcagcagcagcagcagcagcagc 693
DB CAGAAAG 681
622 CAGAAAG 681
694 aaagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 753
DB AAG 741
682 AAG 741
754 gtcctatcccaagagagagagagagagagagagagagagagagagag 813
DB GTCCCTTGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 801
742 GTCCCTTGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 801
814 catgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 873
DB CATGACTTGAG 861
802 CATGACTTGAG 861
874 ggcagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 933
DB GGCAG 921
862 GGCAG 921
934 aagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 993
DB AAGGAG 981
922 AAGGAG 981
994 atgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1053
DB GTTGAAG 1041
982 GTTGAAG 1041
1054 agccacatcatcagcagcagcagcagcagcagcagcagcagcagcagc 1113

DB 1042 TCTTATGCTCCATCGATCATCTTGGATCCAGAGAGAGAGAGAGAG 1101
OY 1114 act--gatalgaagagcagcagcagcagcagcagcagcagcagcagcagc 1170
DB 1102 ACTGAG 1161
OY 1171 attgag 1230
DB 1162 ATCCGAG 1221
OY 1231 gtcgag 1290
DB 1222 ACGAG 1281
OY 1291 acatcttgagagagagagagagagagagagagagagagagagagag 1350
DB 1282 ACCAG 1341
OY 1351 ggaag 1410
DB 1342 GGAAG 1401
OY 1411 aaatctcgcagcagcagcagcagcagcagcagcagcagcagcagc 1456
DB 1402 AAGTCTCCGCAAAATCTGCTGACAGAGAGAGAGAGAGAGAGAG 1447

RESULT 2

US-09-234-613-54
Sequence 54, Application US/09234613
Patent No. 6132973

GENERAL INFORMATION:

APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.

TITLE OF INVENTION: HUMAN REGULATORY MOLECULES

NUMBER OF SEQUENCES: 98

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSO for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/234,613

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/933,750

FILING DATE: September 23, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0356 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 1750 base pairs

RESULT 8
US-08-514-014-7
; Sequence 7, Application US/08514014

```

1  TITLE OF INVENTION: ENCODED THEREBY
2  NUMBER OF SEQUENCES: 12
3  CORRESPONDENCE ADDRESS:
4  ADDRESS: Genetics Institute, Inc. -- Legal Affairs
5  STREET: 87 Cambridgepark Drive
6  CITY: Cambridge
7  STATE: Massachusetts
8  COUNTRY: USA
9  ZIP: 02140
10
11  COMPUTER READABLE FORM:
12
13  MEDIUM TYPE: Floppy disk
14  COMPUTER: IBM PC compatible
15  OPERATING SYSTEM: PC-DOS/MS-DOS
16  SOFTWARE: PatentIn Release #1.0, Version #1.25
17
18  CURRENT APPLICATION DATA:
19  APPLICATION NUMBER: US/08/833,823
20  FILING DATE: 10-APR-1997
21
22  CLASSIFICATION: 530
23
24  PRIOR APPLICATION DATA:
25  APPLICATION NUMBER: 08/514,014
26  FILING DATE: 11-AUG-1995
27
28  ATTORNEY/AGENT INFORMATION:
29  NAME: Brown, Scott A.
30
31  REGISTRATION NUMBER: 32,724
32  REFERENCE/DOCKET NUMBER: G16000
33  TELECOMMUNICATION INFORMATION:
34  TELEPHONE: (617) 498-8224
35  TELEFAX: (617) 876-5851
36
37  INFORMATION FOR SEQ ID NO: 7:
38  SEQUENCE CHARACTERISTICS:
39  LENGTH: 966 base pairs
40  TYPE: nucleic acid
41  STRANDEDNESS: double
42  TOPOLOGY: linear
43  MOLECULE TYPE: cDNA
44  HYPOTHETICAL: NO
45  FEATURE:
46
47  NAME/KEY: CDS
48  LOCATION: 67..348
49
50  OS-08-833-823-7

```

Query Match	2.6%	Score 49	DB 2	Length 966
Best Local Similarity	55.8%	Pred. No. 0.017	Mismatches 91	Conservative 1
				Indels 0
				Gaps 0
QY	1750	cggagttgattgcccgaaccctactcttgcataccatgataagatgaattcttaac	1809	
Db	781	CTCTGTGTATGATCTTTATGCTATATTACATCTGTGTGTACAGTGAACATTTGACAT	840	
QY	1810	gaatgcgaactgcacatgaccttaattcttctaaatgctcaataagcaacgaaatgttc	1869	
Db	841	TATTACTGGAGTCAGGCCCTTATAAGTCAAAAGCACCTATGTGTCGTAAGCATTCCTCA	900	
QY	1870	tacaacmwtlaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	1912	
Db	901	AACATTTAATAAAAAAAAAAAAAAAAAAAAAAAAAA	943	

RESULT 1072B
US-08-104-072B-2
Sequence 2, Application US/08104072B
Patent No. 5639948
GENERAL INFORMATION:
APPLICANT: Michiels, Frank
APPLICANT: Morioaka, Shinji
APPLICANT: Scheirlinck, Trees
APPLICANT: Komari, Toshiko
TITLE OF INVENTION: Stramen-specific Promoters from Rice
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 North 5639948west Center

```

APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1108-A
CURRENT APPLICATION NUMBER: US/09/247,373B
CURRENT FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 08/924,747
PRIOR FILING DATE: 1997-09-05
NUMBER OF SEQ ID NOS: 56
SOFTWARE: Microsoft Office 97
SEQ ID NO 33
LENGTH: 1117
TYPE: DNA
ORGANISM: SOYBEAN
FEATURE:
NAME/KEY: unsure
LOCATION: (1101)
OTHER INFORMATION: M-A OR C
NAME/KEY: unsure
LOCATION: (1104)
OTHER INFORMATION: M-A OR C
NAME/KEY: unsure
LOCATION: (1116)
OTHER INFORMATION: N-G OR A OR T OR C
US-09-247-373B-33

Query Match          2.6%; Score 48.8; DB 4; Length 1117;
Best Local Similarity 73.8%; Pred. No. 0.02; 19; Indels 0; Gaps 0;
Matches 59; Conservative 2; Mismatches

Qy 1833 tatctcattatgcatcaaaagcataacgaactgtttccatacaacmtwlaaaaaaaaaa 1892
    | | | | | | | | | | | | | | | | | | | : | | | | | | | | | |
Db   960 ttatatatgatggagaataaatcatatcatatcattttaaaaaaaaaaaaaaaaaaaaa 1019
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1893 aaaaaaaaaaaaaaaaaaaa 1912
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1020 aaaaaaaaaaaaaaaaaaaa 1039

RESULT 12
US-09-121-321-15
Sequence 15, Application US/09121321
Patent No. 6090783
GENERAL INFORMATION:
APPLICANT: Saiga, Akihiko
APPLICANT: Orta, Satoshi
APPLICANT: Igarashi, Hisanaga
APPLICANT: Okumura, Kouichi
APPLICANT: Sakaguchi, Gaku
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/121,321
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/933,803
FILING DATE: 19-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
```

REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-12C1P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3777 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 139..2151
US-09-121-321-15

Query Match 2.6%; Score 48.8; DB 3; Length 3777;
Best Local Similarity 59.7%; Pred. No. 0.033;
Matches 80; Conservative 1; Mismatches 53; Indels 0; Gaps 0;

QY 1779 gtaaccatgataatgataagattctacagaatgcaactgcaatgcttattttc 1838
DB 3634 GGAATCTTGGACATTAATAAAGGTATCCAGATTCGACACTGCTCAGAGCTGG 3693
QY 1839 taatgtccataaagcatacgaatgttctctcaacmtwtaaaaaaaaaaaaaa 1898
DB 3694 GGGTCTTCACCTTGCCCTACAGTAAAGCCTCAATGAAGTGAAGAAAAA 3753
QY 1899 aaaaaaaaaaaaaa 1912
DB 3754 AAAAAAAAAAAAAA 3767

RESULT 13

US-08-933-803A-15
Sequence 15, Application US/08933803A
Patent No. 6218522
GENERAL INFORMATION:
APPLICANT: Saiga, Akihiko
APPLICANT: Orita, Satoshi
APPLICANT: Igarashi, Hisanaga
APPLICANT: Sakaguchi, Gaku
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,803A
FILING DATE: 19-SEP-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-12C1P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:

LENGTH: 3777 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 139..2151
US-08-933-803A-15

Query Match 2.6%; Score 48.8; DB 4; Length 3777;
Best Local Similarity 59.7%; Pred. No. 0.033;
Matches 80; Conservative 1; Mismatches 53; Indels 0; Gaps 0;

QY 1779 gtaaccatgataatgataagattctacagaatgcaactgcaatgcttattttc 1838
DB 3634 GGAATCTTGGACATTAATAAAGGTATCCAGATTCGACACTGCTCAGAGCTGG 3693
QY 1839 taatgtccataaagcatacgaatgttctctcaacmtwtaaaaaaaaaaaaaa 1898
DB 3694 GGGTCTTCACCTTGCCCTACAGTAAAGCCTCAATGAAGTGAAGAAAAA 3753
QY 1899 aaaaaaaaaaaaaa 1912
DB 3754 AAAAAAAAAAAAAA 3767

RESULT 14

US-08-956-307B-12/c
Sequence 12, Application US/08956307B
Patent No. 6090911
GENERAL INFORMATION:
APPLICANT: Petka, Wendy A.
APPLICANT: Tittel, David A.
APPLICANT: Kevin P. McGrath
TITLE OF INVENTION: REVERSIBLE HYDROGELS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,307B
FILING DATE: 22-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07880/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-956-307B-12

Wed Nov 14 08:34:11 2001

us-09-589-510-3.std.mri

Page 10


```

82 ATGAGATTGAGGAGTGAAGACACTAGCAGACGACGACGATCCCTC 131
17 RHSTHRIHISILEYSGLYLEUGLYLEUASPALAASNGLYMETALALEA 34
132 CCACAGCCACGTGAAGAGGCTGGGCTGAGCAGAGCGCTGGCCCAAGC 181
34 LALEUALAALAGLYPHEVALGlyGlnAlaAlaIaArgGluAlaAGLY 50
182 AGCGGCCCTCAGGGCTTGTGGCCAGAGAACGGCCGAGAGCATGTGGC 231
51 LEUALAVALASPMEIIEARGGLNLYSLEUAGLYAVALALE 67
232 GTCATAGTGAATTAATGAAAGCAAGAAATGGCTGGAAGAGCTGCTT 281
67 ULEUALAGLYPROVALAATHRGlyLysThrAlaLeuAlaLeuGLYIleA 84
282 GTTGGCAGAGACCTCTGCACTGGCAAGACAGCTCTGCTGGCTATTG 331
84 IAGLGLILEUGLYSERLYSVALPROBHECYSPOKEVALGlySERGLU 100
332 CTCAGAGACTGGGTAGTAGGTCCCTTCTGCCAATGTGGGAGTAGAA 381
101 VALTYRSERSEGLUVALLYSlyThrGluValLeuMetGluAspPheAR 117
382 GTTACTCACTGAGATCAAGAGACAGAGCTCTCATGACAGACTTCG 431
117 GARGALALEGLYLEUARGILEYSGLYASNLYSGLUVALTYRGlyGlyG 134
432 CAGGGCCATTGGCTGGGATTAAGAGACCAAGAAAGTTATGAAGTG 481
134 LUVALTHRGlyLeuSerProGlyGluAlaGlySerThrThrGlyGlyYr 150
482 AAGTCACAGAGCTAACCTCCGTGAGACAGAGAACATCCATGGAGAT 531
151 ALALYSERILESERHISVALILELESERILEUYSlyThrValLYSGlyTh 167
532 GGCAAAACCATTAAGCATGTGATCATAGACTCAAAACAGCCAAAGAAC 581
167 ILYSGlyLeuLYSLeuASPSErSeriLeTYRASPALAILEUILEYSGlyL 184
582 CAACAGCTTGAAGAACTGACCCAGCATTTTGAAGATTTGAGAAAGAGC 631
184 YSVALALEVALGlyAspValIleTYRILEGLUALAASERGLYALALEA 200
632 GAGTGAAGACCTGGAGATGTGATTTACATTAGACCAACAGTGGGCCG 681
201 LYSARGVALGlyArgCysASPSErPheAlaThrGlyTYRASPLeuGluA 217
682 AAGAGCGAGGGCAGGTGTGATACCTATGCCACGAATTCGACCTGTAAGC 731
217 AGLGLIUTYRVALPROILEPROLYSGlyGluValHISLYSlySGLU 234
732 TGAAGAGATGTCCCTTCCAAAGGGAGTGTGCCAAAGAAAGAAAGAA 781
234 LEVALGlnASPVALTHrLeuHISASPLeuASPALAAlaAsnLysValAla 250
782 TCATCCCAAGATGTGACCTTGATGACTTGATGGCTAAATGCGGGCC 831
251 GlnGlyLYGlnASPLeuSerLeuMetGlyGlnMetLysProAR 267
832 CAGGGGGGCAAGATATCTCTGTCCATGTAGGGCCAGCTAATGAAGCCAA 881
267 GLYSThrGlyIleThrGlyLysLeuArgGlnGlyLeuSlySValValA 284
882 GAAGACAGAAATCACAGAAACTCGAGGGGACATTAATAGGTGTGA 931
284 SPARGTYRILESPGLUGlyIleAlaGlyLeuValProGlyValLeuPhe 300
932 ACAAGATCACTGACCGGCAATGTGTGAGCTGCCGGGTGTGCTGTT 981
301 ILEASPGlyUValHISMetLeuAspIleGlyCysPheSerTYRLeuSnr 317
982 GTTGATGAGGTCCACATCTGACATTTGAGTGTCTTCACTACTGACCG 1031

```

```

317 GAlaLeuGlySerProLeuSerProIleValIleLeuAlaThrAsnArg 334
1032 CGCCCTGGAGTCTTCTATCGCTCCATCGTCATCTTTGTATCAACCGAG 1081
334 LYIIECYSASNVALARGlyThr...ASPMEIIErSeriProHISGLYIle 349
1082 GCAACTGTGATCAAGAGCGACTGAGACATCATCAATCCCTCAGCGCATC 1131
350 PROVALASPLeuLeuASPArgLeuValIleIleArgThrGlyThrGly 366
1132 CCHTGTGACCTTCTGACCGAGTGAATGATGATCGGAGCATGCTATAC 1181
366 YPROTHRGlyMetIleGlnIleLeuAlaIleArgAlaGlnValGlyL 383
1182 TCCACAGGAAATGAACACAGATCATTAATCCGTGCCAGAGGAAAGAA 1231
383 LEASPMEIASPGlyLeuSerLeuAlaTYRILEYSGlyGlyGln 399
1232 TCACATCACTGAGGAGGACTGACCAACCTGGGGGAGATTGGCACCAAG 1281
400 THRSErLeuArgHISAlaIleGlnLeuIleSerProAlaSerValAla 416
1282 ACCACACTGAGGTACTCACTGACCTGACCCCGGCCAATCTGCTGTC 1331
416 ILYSThrASNGLYArgGlyLysIIECYSlySAlaASPLeuGlyGlyValS 433
1332 TAAATCAACGGAAGAGACACATTTGANAAGAGCATGTGCAAGAGATCA 1381
433 ERGLYLeuTYRLeuASPAlaLYSSErSeriAlaArgLeuLeuGlnGly 449
1382 TGAACTTTCTATGATGATGCCAAGTCTCCGCCAAATCTGTGCTGACCG 1431
450 GlnGluArgTYRILE 454
1432 CAGGATTAAGTACATG 1446

seq_name: /g9n2_6/plodata/2/ina/6a_COMB.seq:us-09-234-613-54
seq_documentation_block:
? Sequence 54, Application US/09234613
? Patent No. 6132973
? GENERAL INFORMATION:
? APPLICANT: Lal, Preeti
? APPLICANT: Hillman, Jennifer L.
? APPLICANT: Bandman, Olga
? APPLICANT: Shah, Purvi
? APPLICANT: Au-Young, Janice
? APPLICANT: Yue, Henry
? APPLICANT: Guegler, Karl J.
? APPLICANT: Corley, Neil C.
? TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
? NUMBER OF SEQUENCES: 98
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Inocyte Pharmaceuticals, Inc.
? STREET: 3174 Porter Drive
? CITY: Palo Alto
? STATE: CA
? COUNTRY: USA
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/234, 613
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/933, 750
? FILING DATE: September 23, 1997
? ATTORNEY/AGENT INFORMATION:

```

Wed Nov 14 08:34:16 2001

us-09-589-510-4.std.in1

Page 3

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 1750 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HMC1N0T01
CLONE: 9476
US-09-234-613-54

alignment scores:
Quality: 1721.50 Length: 455
Ratio: 4.109 Gaps: 1
Percent Similarity: 92.088 Percent Identity: 72.527

alignment block:
US-09-589-510-4 x US-09-234-613-54 ..

Align seg 1/1 to: US-09-234-613-54 from: 1 to: 1750

1 MetArgIleGluGluValAlaGlnSerThrSerIleuGlnIleAlaIle 17
|||||
82 ATGAGCATTTGAGGAGGTGAAGACACACACAGACGACGCGATCCCTC 131
17 rhtsthrstleuGlyLeuGlyLeuAspAlaAsnGlyMetAlaIle 34
|||||
132 CCACAGCCACGTGAAGGCTGGGCTGGACGAGACGCGCTGGCCAGC 181
34 leuAlaIleGlyPheValGlyGlnAlaAlaAlaArgGluAlaIle 50
|||||
182 AGCGGCGCTCAGGCGCTGTGGCCAGGAGAACGCGCGAGGCGATCGGC 231
51 leuAlaValAspMetIleArgGlnIleuGlyMetAlaGlyAlaValle 67
|||||
232 GTCTAGTACATTAATTCAGAAAGCAAAATGCTGGAGAGCTGCTT 281
67 uLeuAlaGlyProProAlaThrGlyIleuThrAlaLeuAlaLeuGlyIle 84
|||||
282 GTTGGCAGGACCTCTGGAAGCTGGACAGACCTCTGCTGCTGCTATTG 331
84 lagGlnIleuGlySerIleValProPheCysProMetValGlySerGlu 100
|||||
332 CTCAGAGACTGGTATGATAGCTCCCTCTGCCCATTGGTGGGAGTGA 381
101 ValTyrSerSerGluValAlaValIleuGluValIleuMetGluAsnPheAr 117
|||||
382 GTTACTACTAGATCAGATCAGACAGACAGAGCTGCTGATGAGAACTCCG 431
117 gArgAlaIleGlyLeuArgIleuGlyLeuGlyLeuGlyValIleuGlyG 134
|||||
432 CAGGCGCATTTGGCTGCAATTAAGGAGACCAAGAGGTTTATGAGAGTG 481
134 luValThrGluLeuSerProGluGluAlaGlnSerThrThrGlyGlyTyr 150
|||||
482 AAGTTCACAGAGCTAAGCTGCTGTGAGACAGAGATCCCATGGAGGATAT 531
151 AlaTyrSerIleSerHisValIleIleSerLeuIleuThrValIleuGly 167
|||||
532 GGCAGAAACCTTATGATGATGATGATGATGATGATGATGATGATGAT 581
167 PysGlnIleuLeuAspSerSerIleuAspAlaLeuIleuGlyGlu 184
|||||
582 CAAAGCTGTAAGTGAAGCTGTAAGCTGTAAGCTGTAAGCTGTAAGCT 631

184 ysvAlaIleValIleGlyAspValIleTyrIleGluAlaAsnSerGluAlaVal 200
|||||
632 GAGTACAGCTGAGATGATGATGATGATGATGATGATGATGATGATGAT 681
201 LysArgValGlyArgCysAspSerPheAlaThrGluTyrIleuGluAla 217
|||||
682 AAGAGCAGGAGGAGTGTATGATGATGATGATGATGATGATGATGATGAT 731
217 agIleuTyrValProIleProGlyGlyValAlaHisIleuGlyGlyIle 234
|||||
732 TGAAGATATGCTCCCTGCAAAAGGAGTGTGCAAAAGGAGGAGGAGGAG 781
234 leValGlnAspValThrIleuHisAspLeuAspAlaAlaAsnAlaGlnPro 250
|||||
782 TCATCCAGATGTGACCTTCCATGATGATGATGATGATGATGATGATGAT 831
251 GlnIleGlyGlnAspIleLeuSerIleuMetGlyGlnMetMetIleProAr 267
|||||
832 CAGGCGGAGCAAGATATCTGTCATGATGATGATGATGATGATGATGAT 881
267 GlyThrGluIleThrGluIleuArgGlnIleuIleAsnIleValAla 284
|||||
882 GAGACAGAAATTCACAGACAACTTGAGGAGGAGATTAATTAAGTGTGA 931
284 snArgTyrIleAspGluGlyIleAlaGluIleuValProGlyValIleuPhe 300
|||||
932 ACAAGTACATGACACGAGGATGCTGAGTGTGCTGCTGCTGCTGCTGCT 981
301 IleAspGluValHisMetLeuAspIleGluCysPheSerTyrIleuAsnAr 317
|||||
982 GTTGATGAGGCTCCACATGCTGACATGATGATGATGATGATGATGATGAT 1031
317 gaIleuGlnSerProLeuSerProIleValIleuAlaThrAsnArg 334
|||||
1032 CGCCCTGAGGCTTCTATGCTCCATGCTGCTGCTGCTGCTGCTGCTGCT 1081
334 IyIleCysAsnValArgGlyThr...AspMetThrSerProHisGlyIle 349
|||||
1082 GCACGTGTGTCATCAGAGGACACAGACATCAGATCAGATCAGATCAGAT 1131
350 ProValAspLeuLeuAspArgIleValIleIleArgThrGluThrTyrG 366
|||||
1132 CCTCTGACCTTCTGACCGAGTATGATGATGATGATGATGATGATGATGAT 1181
366 yProThrGluMetIleGlnIleuAlaIleArgAlaGlnValGluIle 383
|||||
1182 TCACAGAGAAATGAACAGATCATTAATTCGCTGCGCCAGAGGAGGAA 1231
383 leAspMetAspGluGlnSerIleuAlaTyrIleuGlyGlyIleGlyGln 399
|||||
1232 TCACATCAGTACGAGGAGGACCTGAACACACCTGGGAGATGTGGCAGCAG 1281
400 ThrSerIleuArgHisAlaIleGlnIleuIleSerProAlaSerValIle 416
|||||
1282 ACCACACTGAGGTACTCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1331
416 rIystrAsnGlyArgGluIleIleCysIleAlaAspLeuGluIleVal 433
|||||
1332 TAAATCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1381
433 erGlyLeuTyrIleuAspAlaIleSerSerIleuAlaArgLeuGlnGln 449
|||||
1382 GTGAACTTCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1431
450 GlnIleuArgTyrIle 454
|||||
1432 CAGGATTAAGTACATG 1446
seq. name: /cgn2_6/plodata/2/ina/6B.COMB.seq: US-08-998-416-1005
seq. documentation_block:
; Sequence 1005, Application US/08998416
; Patent No. 6239264

228 GCAGGCGTGATTCTCAAAATGGTGCAGAACGGTACGATTGCCGGCGGCGC 179

178 CGCGTGGTCTGGTCCGCCGTGACCGCGCAAAACGGCGCTGGCGATGC 129

128 GCTTGTCCAGTCACTGGGCGCAGATGTGCCCTTCACAGCCATCCGCGCG 98

99 SerGluValTyrSerSerGluValAllyLysLysThrGluValLeuMetGluAs 115
|||||:::|||||||.....

115 GPheArdaArdAlaIecGwTwtTtTtt
160 TctcGcgTGGAGCTATCCAGAACCAGGCCCTTGACTCAGGC 29

28 GTTCGGAAGTCGATTGCTGTCAGATC 1

name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1

Sequence 1, Application US/09103840A
 Invent No. 62994328

APPLICANT: FLEISCHMAN, Robert D.

PLICANT: FRASER, Claire M.
PLICANT: VENTER, John C.

TITLE OF INVENTION: SEQUENCES FOR STRAIN ANALYSIS IN MYCOPLASMA TUBERCULOSIS
 PRIORITY REFERENCE: 24366-20007.00

RENT FILING DATE: 1998-06-24
 NUMBER OF SEO ID NOS.: 2

```
NAME: Patenlin Ver. 2.1
ID NO 1
NGTH: 4411530
```

PE: DNA
ORGANISM: Mycobacterium tuberculosis

103-840A-1

```

nt_scores:
  Qualitv. 131 00

```

Ratio:	0.582	Gaps:	29
t Similarity:	44.643	Percent Identity:	22.024

nt_block:
589-510-4 x US-09-103-840A-1/rev

seg 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411

[illegible]

7 rHISThrnIleLysGly.....[euNG]vLeuAsnA 38
IIIIII

3GCACGGTCGTCACACGAGGCATCCTGGGGGAGCTGCTGCTCTACGTG 4052524

[illegible]

.....AlaAlaArgGluAlaAlaGlyLeuAlaValAspMetIle.Argg 58
|||
|||

22CCGATGCAAGCGCGCGCCCGATGGCTTCGGGTTCCGCAAGTCACGCG 4052433

.....Ala 62
|||:::
|||:::
GCCAAGCACTGAGCAGGACATGCCCAAGACCCACCTTCCCCCG

```

63 GLY.....ArgAlaValL 67
4052382GCTTCAGCAGGCGGTGAGAGCTCTACGATCATCAGGACTTCCTGCA 4052333
64 .....
4052323GACCCAGCAGTACCAAGCGCTGGGCGCCAGATCCCAAGAGGGTGC 4052283
67 eulEuAlaIaglyProProAlaThrhGlyThrAlaLeuAlaLeuAlaLeuGlyLe 83
4052282TCCTTACGGGCGCGGGAACGGTAGACGTTGCTGCTGCTGCTGGGTC 4052233
84 AlaIngIuLeuGlySerIysValProPheCysProMetValGlySerG 100
4052232GCGCGGCAAGCCGA.....GTCCGTTCTTCACCATCTCCGCGTCCGA 4052189
100 u.....ValTYrSerSerGluValAllyLysThrG 110
4052188CTTCGTGCAATGTTCTGCGGCGGCATCCCGTGTGCA..... 4052146
110 LuValLeuMetGluAsnPheArGArGAlaIleGlyLeuArGlyIleYsglu 126
4052145.....GACCTGTCGAGCAGCGC.....AACGAG 4052122
127 AsnYsglu.....ValTYrGluGlyGluValThrhGluLeuSerProG 141
4052121AACACCGCGTGATCTTCGTGAGAGATCGACGCGCGCGCGAGCA 4052072
141 uGluAlaGluSerThrhThrhGlyTYrAlaLysSerIleSerIleValI 158
4052071AAGAGGCGCGGCGGTGCGGCGGTGACGACGCT..... 4052035
158 leIleSerIeuYsrThrValIlySgIyThrIlySgIuLeuYsLeuAspSer 174
4052034.....GAG 4052032
175 SerIleTYrAspAlaLeuIleLysGluYsValAlaValGlyAsp..... 189
4052031CAGACCCCTCAACAGTGTCTACTGCAATGAGACGGTTTGGCATCCGCGC 4051982
190 ....ValIleTYrIleGluAlaAsnSer.....GlyA 199
4051981CGCGGCTACTGCTGATCGCGGCGCCACCAACCGCGCCATCTCGACCGCG 4051932
199 laValIysArGValGlyArGysAspSerPheAlaThrhGluTYrAspLeu 215
4051931CGCTGTGGCGCGCGCGCTTCGAC..... 4051906
216 GluAlaGluGluTYrValProIleProIlySgIyGlu..... 227
4051905.....CGCCAGATCCCGGTATCCACCCCGATCTGGCGGCGTCCGCG 4051865
228 .....ValIisLysIys...LysGluIleValGlnAspValT 239
4051864GGGCGTCTCGGCTGCTCAACAGGGAAGCGCATGGCGCGGAGCGCC. 4051816
239 hrLeuHisAspLeuAspAlaAlaAsnAlaGlnProGlnIy.....Gly 253
4051815.....GACCTTCGACGCGCTGCGCAAGCGGACGCTGCGCATGACCGGA 4051774
254 GluAspIleLeuSerLeuMetGlyGln.....MetMetLysProArGly 268
4051773GCGACCTGGCCACATCATCAACAGGCGGCGCTCTGACCGCGCGGA 4051724
268 s.....ThrGluIleThr...GluIysIeuArGInGluIleAsnIysV 282
4051723GACAGCGACCGTATCACCGCTCCGCTCGAGGAAGCGGTGGACGGG 4051674
282 aIVal.....AsnArGTYrIleAspGluGlyIleAla 292
4051673ATGATCGGCGCGCGCGCGCAAGGCGGATCTCATCAGCAGCAGGAGAG 4051624
293 GluLeuValProGlyValLeuPheIleAspGluValHisMetLeu..... 307

```

```

4051623AAGATC.....ACCGCTATCAAGAGGGGCGGCACACCTGCGCGC 4051583
308 .....AspIleGluCysPheSerTYrIleuAsnArGAlaLeuG 320
4051582TTGGCGATGCGCGATATCGAG..... 4051561
320 LuSerProLeuSerProIleValIleLeuAlaThrhAsnArGlyIleCys 336
4051560.....CCGATTATTAAGTACAGATCTGCGCGCGCGGCT..... 4051525
337 AsnValArgIyThrhAspMetThrhSerProHisGlyIleProValAspLe 353
4051524.....ACCGGCGGCGACGCGGTGCGTCCGGA 4051496
353 uLeuAspArgLeuValIleIleArGThrhGluThrTYrGlyProThrhGlu 370
4051495AAGAACAAAGCGCTGCGGAGACCCGCTCGAATGATC.....GCCGAC 4051452
370 eTleGInIleLeuAlaIleArGAlaGlnValGluGluIleAspMetAsp 386
4051451TGCTTTCGATGGGTGGCGCGCC...GCCGAAAGACTGTTGTTGTT 4051405
387 GluGluSerLeuAlaTYrLeuGlyGluIleGlyGlnGlnThrhSerLeuAr 403
4051404GAGCGGACACCGCGCGGTGTCGACATCGAGCAGCGCCACCAAGATAGC 4051355
403 gHisAlaIle 406
4051354GCGCTCAATG 4051345

```

seq_name: /cgn2_6/prodata/2/lna/6h_COMB.seq:us-09-103-840A-2

```

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

alignment_scores:
Quality: 130.50 Length: 400
Ratio: 0.690 Caps: 17
Percent Similarity: 47.250 Percent Identity: 22.000

```

alignment_block:

US-09-589-510-4 x US-09-103-840A-2/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

```

25 GlyLeuAspAlaAsnGlyMetAlaIleAlaLeuAlaIleGlyPheValG 41
|||||
2875941GCTGCGATGCTGCGCGCT.....GCTGAGCAAGATGG 2875907

```


Quality: 117.00 Length: 219
Ratio: 1.045 Gaps: 13
Percent Similarity: 51.142 Percent Identity: 25.114

alignment block:

US-09-589-510-4 x US-08-999-774A-5 ..

Align seg 1/1 to: US-08-999-774A-5 from: 1 to: 2932

```

64 ArgValValLeuLeuAlaGlyProProAlaThrGlySerThrAlaLeuAl 80
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
980 AGAGATTCCTGCTTACGACCTCCGAGACAGAGAAAGCCCTGATTCG 1029
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
80 AlaGlyAlaLeuGlnGluLeuGlySerLysValProPheCysPromely 97
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1030 TCGAGCTGTAAGTAATGACACTGAGCCTTC.....TTCTTCTGATCA 1073
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
97 aLgYserGluValTyrSerSerGluValLysLysThrGluValLeuMet 113
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1074 ATGCTCTGAGATCATGAGCAAAATGGCTGTGATGTCGAG..... 1114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
114 GluAsnPheArgArgAlaLeuArgLleLysGluAsnLysGlu.. 129
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1115 AGCAACCTTCTGTAAGCCTT.....GAGGAGGCTGAGAA 1149
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
130 .....ValTyrGluGlyGluValThrGluLeuSerProG 141
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1150 GAATGCTCTGCCATCATCTTCATGTGATGAGTAGATGCATGCTCCCA 1199
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
141 LngLValGluSerThrThrGlyLysThrLysSerLleSerHisVal 157
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1200 AAAAAA...GAGAAACTCATGCGAGGTGAGCGCGCCATGTATCATCAG 1246
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
158 lleleSerLeuLysThrValLysGlyThrLysGlnLeuLysLeuAspSe 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1247 TTGTGTAACCTC.....ATGATGGCCTTAAGCAGAG..... 1279
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 rSerlleTyrAspAlaLeuLleLysGluLysValAlaValGlyAspVal 191
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1280 .....GCACATGTGA 1289
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
191 leTyrleGluAlaAsnSer.....GlyAlaValLys 201
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1290 TTGTATGCGACGACCAACAGACCACACCATTTGACCCGCTCTACGG 1339
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
202 ArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGluAlaG 218
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1340 CGATTGTGTCG.....TTTGACAGGAGGTAGA 1368
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 uGluTyrValProLleProLysGly.....GluValHis. 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1369 TATTGGAATTCCTGATGCTACAGGACGCTTAGAGATTCTTCAGATCCATA 1418
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
230 ..LysLysLysGluLleValGlnAspValThrLeuHisAspLeuAspAla 245
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1419 CCAAGAAATGAAAGCTGCGACATGATG.....GACCTGGAACAG 1459
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 AlaAsnAlaGlnProGlnGly.....GlyGlnAspLleLeuSerLeu 260
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1460 GTAGCAATGAGACTCAGCGGCGATGGTGTGCTGACTTAGACGCCCTGTG 1509
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
260 tGlyGln 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1510 CTCAGAG 1516
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:US-09-446-504-4

seq_documentation_block:

```

; Sequence 4, Application US/09446504
; Patient NO. 6218150
; GENERAL INFORMATION:
; APPLICANT: DEMORI, Takashi
; APPLICANT: SATO, Yoshimi

```

```

; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazuo
; APPLICANT: MURAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/446,504
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
; US-09-446-504-4

```

alignment_scores:
Quality: 111.50 Length: 367
Ratio: 0.664 Gaps: 18
Percent Similarity: 45.777 Percent Identity: 21.526

alignment block:

US-09-589-510-4 x US-09-446-504-4 ..

Align seg 1/1 to: US-09-446-504-4 from: 1 to: 984

```

54 AspMetlleArgGlnLysLysMetAlaGlyArgAla..... 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
76 GACATTGTGAGACAGACGACATAGTGAAGCAAGCTCAACACTAGCTCAA 125
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
66 .....ValleuLeuAlaGlyProProAlaThrGlyL 76
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 AACGTGATCAATGCCCCACTACTCTTCGAGAGCCCGCTGTGCGAA 175
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
76 ysrAlaLeuAlaLeuGlylleAlaGlnLleuGlySerLysValPro 92
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
176 AGACTACAGCGCTTTGGCCCTTGCAAGAGAGCTT..... 210
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 PheCysPromelyAlaGlySer.....GluValTyr 102
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
211 .....TTGCGCGAAGAACTGGAGGCATTAACCTTCCTGAGTTGAA 248
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 rSerSerGluValLysLysThrGluValLeuMetGlu.....Asp 116
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
249 TGCTTCAGATGAAGAGGTATTAACGTATTTACAGAGAAAGTTAAGAGAT 298
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 heArgArgAlaLleGlyLeuArgLleLysGluAsnLysGluValTyrGlu 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
299 TTGCGAGAACAAAGCCTTATAGAGAGCAAGCTTCAAGATTAATTTCCCT 348
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 GlyGluValThrGluLeuSerProGlnGluAlaGluSer.....Th 146
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
349 GATGAGCGCGCGCTTAACTCAAGATGCCCAACAGCCCTTAAGAGAAC 398
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 rThrGlyGlyTyrAlaLysSerLleSerHisValLleLleSerLeuLysT 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
399 CATGAAATGTTCTCGATACCTTCGCTTTATCTTGAGCTGTAAGTACT 448
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 hrValLysGlyThrLysGlnLeuLysLeuAspSerSerlleTyr..... 177
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
449 CTCGCAAGATTAATGAACCATACAGCTAGATGTCAATATTCGCTTC 498
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
178 .....AspAlaLeuLleLysGluLysValAlaValGlyAspVal 190
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
499 AGACTCTCCGCGATGAGATTAAGCGAAGAGACTAAG..... 537
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

190 111eTrrllegluAlaasnserglYalValVArgValGlyArgCysA 207
      ||||| ||||| |||||
538 ....TACATTGCCGAAATAGAGGCTTAGAG..... 564
207 spSerPheAlaThrGluTyrAspLeuGluValGluTyrValProIle 223
      ||||| ||||| |||||
565 .....CTAAGTGAAGAAGGTCTCCAAACATCTTTC.....ATA 600
224 ProlGlyGluValHisLysLysLysGluIleValGlnAspValThrIle 240
      ||||| ||||| |||||
601 GCAGGAAGATATGAGAGCAATTAACATTCTGCAAGCTTCAGACAGC 650
240 uHisAsp.....LeuAspAlaAlaAsnA 248
      ||||| ||||| |||||
651 TCTAGACAGAGATCACCAGCAAAACCTATTTCATGTCGAGTCAGTAGAG 700
248 1aGlnProGlnGlyGlnAsp.....IleLeuSerLeuMetGlyGln 262
      ||||| ||||| |||||
701 CTAGACCTGAAGATATAGAGAGATGATGCTTCTCTCTCAAGGCAAC 750
263 MetMetLysProArgLysThrGluIleThrGluLysLeuArgGlnGluI 279
      ||||| ||||| |||||
751 TCTTGAAGGCCAGA.....GAAAGCTTAGGAGATGACT 785
279 eAsnLysValAlaAsnArgTyrIleAspGluIleAlaGluLeuValP 296
      ||||| ||||| |||||
786 TCTCAAG.....CAAGGACTTACT..... 804
296 roGlyValLeuPheIleAspGluValHisMetLeuAspIleGlyCysPhe 312
      ||||| ||||| |||||
805 ..GGAGACATGTACTAGTTCACATCCACAAA.....GAGCTTTC 843
313 SerTyrLeuAsnArgAlaLeuGlnSerProLeuSerProIleValIleLe 329
      ||||| ||||| |||||
844 .....AACCTGCCAATAGAGGCCAAGAG.....GTCTGCT 878
329 uAlaThrAsnArgGlyIleCysAsnValArgGlyThrAspMetThrSerP 346
      ||||| ||||| |||||
879 TGCTGATTAAGATAGAGATATTAATCTCAGACTCGTTGAA..... 918
346 roHisGlyIleProValAspLeuAspArgLeuValIleArgThr 362
      ||||| ||||| |||||
918 ..... 918
363 GluThrTyrGlyProThrGluMetIleGlnIleLeuAlaIleArgAlaG 379
      ||||| ||||| |||||
919 .....GGGCTTAATGAATAATATCTAGCTTGAAGACATCTTAGCAC 959
379 n 379
      ||||| ||||| |||||
960 G 960
seq_name: /cgn2_6/plodata/2/ina/6A_COMB.seq:us-08-987-123-1
seq_documentation_block:
; Sequence 1, Application US/08987123C
; Patent No. 6136557
; GENERAL INFORMATION:
; APPLICANT: Eli Lilly and Company
; FILE OF INVENTION: STREPTOCOCCUS PNEUMONIAE GENE SEQUENCE ftsH
; PRIORITY REFERENCE: X-11762 Sequence Listing
; Patent No. 6136557
; CURRENT APPLICATION NUMBER: US/08/987,123C
; CURRENT FILING DATE: 1997-12-08
; PRIOR APPLICATION NUMBER: 60/036,281
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1956)
US-08-987-123-1

```

```

Alignment_scores:
  Quality: 111.50      Length: 484
  Ratio: 0.474         Gaps: 31
  Percent Similarity: 48.554      Percent Identity: 23.140

```

```

alignment_block:
US-09-589-510-4 x US-08-987-123-1

```

```

Align seg 1/1 to: US-08-987-123-1 from: 1 to: 1959

```

```

39 PheValGlyGlnAlaAlaArgGluAlaGluLeuAlaValAspPhe 55
      ||||| ||||| |||||
556 TTTTCAGATGTTGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGT 605
55 tIle.....ArgGlnLysMetAlaGlyArg 64
      ||||| ||||| |||||
606 TGTTCAGTCTTAAAGATCCAAACGATTCACAAACCTTGAGCCGTA 655
65 .....AlaValLeuLeuAlaGlyProProAlaThrGlyLysThrAla 78
      ||||| ||||| |||||
656 TTCCAGCAGGTGTTCTTTGGAGGAGACCTCCGGGACAGGTAAACCTTG 705
79 LeuAlaLeuGlyIleAlaGlnGluLeuGlySerLysValProPheCysPr 95
      ||||| ||||| |||||
706 CTTCGTAAAGGAGCTGCTGAGAGAGAGAGT.....GTTCATCTTCTTAG 749
95 oMetValGlySerGlu.....ValTyrSerSerG 105
      ||||| ||||| |||||
750 TATCTCAGGTTCTTACCTTTAGAAATGTTTGTCCGAGTTGAGCTAGTC 799
105 LuValLysLysThrGluValLeuMetGluAsnLysGluValTyrGluGly 119
      ||||| ||||| |||||
800 GTGTTCGCTCT.....CTTTTGGAGATGCGCAAAAGCAGCACCA 840
120 .....IleGlyLeuArgIleLysGluAsnLysGluValTyrGluGly 134
      ||||| ||||| |||||
841 GCTATCATCTTTATCGATCTAAATGATGCTGTGAGACGTCACAGCTGAGT 890
134 uValThrGluLeuSerProGluGluAlaGlnSerThrThrGlyGlyTyrA 151
      ||||| ||||| |||||
891 CGGCTCGGGGAGAGTAATGACGACGTAACAAC..... 927
151 1aLysSerIleSerHisValIleIleSerLeuLysThrValLysGlyThr 167
      ||||| ||||| |||||
928 .....TTGAACCAACTTTTGTATGATGATGATGATGATGATGATGAT 969
168 LysGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGluY 184
      ||||| ||||| |||||
970 GAA..... 972
184 sValAlaValGlyAspValIleTyrIleGluAlaAsnserGly..... 198
      ||||| ||||| |||||
973 .....GGGATTATGCTCATGCCCTGGAGCAAAACGCTTCAGATGTAC 1012
199 .....AlaValLysArgValGlyArgCysAspSer.....PheAla 210
      ||||| ||||| |||||
1013 TTGATCCTGCCCTTTTGCCTCCAGAGCTTTGATGATGAAGAAATATGCT 1062
211 ThrGluTyrAspLeuGluAlaGlnGluTyrValProIleProlGlyGly 227
      ||||| ||||| |||||
1063 GGCCTGCTGATGTTAAGGTCGTGAAGCAATCTTG.....AA 1100
227 uValHis...LysLysLysGluIleValGlnAspValThrLeuHisAsp 243
      ||||| ||||| |||||
1101 AGTTACGCTAAGAACACAGCCTTTACGACAGATGTT.....GATT 1141
243 euAspAlaAlaAsnAlaGlnProGlnGly.....GlyGlnAspIleLeu 257

```



```

1142 TGAATATAGAGCTCAACAACTCCAGCTTGTGCTGATTTAGAG 1191
      :::: ||| ||| ||:::
258 SerLeuMetGlyLys.....MetMetLysProArgLysThrGluLeuH 272
      ::::: ||:::
1192 AATGCTTGAATGAAACACCTTAGTGTGCTGCTGCTGC..... 1230
272 rglLysLeuArgGlnGluLLeuAsnLysValVal.....AsnArgTyr 287
      ||||| ::::
1231 .....AATAATCGATTAATGATGCTTCAGATA 1258
287 LeaspGluLysLLeuLagLLeuValProGlyValLeuPheLLeaspGlu 303
      ||||| ::::
1259 TTGATGACAGACAGATGAGTTATGCTGACCTTCAAGAAAGATAG 1308
304 ...ValHisMetLeuAspLLeuLysPheSerTyrLeuAsnArgAlaLe 319
      ||| ::::
1309 ACAGTTTCACAAAAGACGAGATTTGTTGCTTAC...CATGAGCGCAGG 1355
319 uGluSerProLeuSerProLLeuValLLeuValLThrAsnArgGlyLLeu 336
      ::::: ||:::
1356 ACATACCATTTGTTGTT.....CTGATCTTGTCTGCTGCTGCTGCTGCT 1399
336 yAsnVal.....ArgGlyThrAspMetThrSerProHisGly 348
      ::::: |||||
1400 ATTAGGTTACAAATTGACACGCGCGCTGACAGCGGATACATGATGCA 1449
349 IlePro.....ValAspLeuLeuAspArg 356
      ::::: |||||
1450 CTTCCTCAAGAGATCAATGCTTCTATCTAAGAGATATGAAAGACCA 1499
356 gLeu.....ValIleLeuArgT 362
      |||
1500 ATTGGCTGGCTTAATGGTGGACGTGATGACGAAATTAATCTTAATG 1549
362 hrcGluThrTyrGly.....ProThrGluMetIle 371
      ::::: |||||
1550 TCCAACACTACAGGACCTTCAACAGCTTGAACAGACGACCAAAATG... 1596
372 GlnIleLeuValLLeuArgAlaGlnValGluLLeuLLeuAspMetAspGlu 388
      ||||| |||||
1597 .....GCACGTGCATAGTGTACAGAGTACGATGAGAGGAAA 1634
388 uSerLeuAlaTyrLeuGlyLysLLeuGlnThrSerLeuArgHisA 405
      ::::: |||||
1635 A.....CTTGGCCCAAGTA.....CAATATGACGAAACACATG 1666
405 La...IleGlnLeuLLeuSerProAlaSerValSerLysThrAsnGly 420
      ||| ::::
1667 CTATGCTTGGTGCACAGAGCTCTCAAAAATTCAGAAACAAACACCT 1716
421 ArgGluLysLLeuLysAlaAspLeuGlnLysLLeuLLeuSerGlyLys 437
      ||| |||||
1717 TATGAA.....ATTGAT...GAAGAGTGTCTTATTATTAA 1751
437 uAspAlaLysSer...SerAlaArgLeuLeuGlnGlnGlnLysArgT 453
      ::::: ||:::
1752 TGAGCGCAGCAATAAAGCTGCTGAATATATTCAGTCAAAATCGTAAACTC 1801
453 yr 453
      ::
1802 AC 1803
seq_name: /cgn2_6/prodata/2/lna/6A_COMB.seq:us-08-987-123-3
seq_documentation_block:
; Sequence 3: Application US/08987123C
; Patent No. 6136557
; GENERAL INFORMATION:
; APPLICANT: Eli Lilly and Company
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE GENE SEQUENCE FISH
; FILE REFERENCE: X-11762 Sequence Listing
; Patent No. 6136557

```

```

; CURRENT APPLICATION NUMBER: US/08/987,123C
; CURRENT FILING DATE: 1997-12-08
; PRIOR APPLICATION NUMBER: 60/036,281
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1959
; TYPE: RNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: mRNA
; LOCATION: (1)..(1959)
US-08-987-123-3

```

```

alignment_scores:
  Quality: 111.50      Length: 484
  Ratio: 0.474         Gaps: 31
  Percent Similarity: 48.554  Percent Identity: 23.140

```

```

alignment_block:
US-09-589-510-4 x US-08-987-123-3

```

```

Align seg 1/1 to: US-08-987-123-3 from: 1 to: 1959

```

```

39 PheValIGlyGlnAlaAlaArgGluAlaLagLLeuAlaValAspMe 55
      ||| ::::: |||||
556 UUUUACAGAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 605
55 tile.....ArgGlnLysLysMetAlaGlyArg. 64
      ::::: |||||
606 UUUUACAGAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 655
65 tile.....AlaValLeuLeuAlaGlyProProAlaThrGlyLysThrAla 78
      ::::: |||||
656 UUUUACAGAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 705
79 LeuAlaLeuGlyLLeuLagLLeuGlySerLysValProPheCysPr 95
      ||||| ::::: |||||
706 CUUUCUUAAGCGAGUGCGUGAGAGAGAGU.....GUUUCUUAUUUUAG 749
95 ometValGlySerGlu.....ValTyrSerSerG 105
      ::::: |||||
750 UUUUACAGAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 799
105 LuValLysLysThrGluValLeuMetGlnAsnPheArgAla..... 119
      ::::: |||||
800 GUUUGUCCUUUU.....CUUUUUUAGAGAGUCCAAAAGACGACCA 840
120 .....IleGlyLeuArgLLeuLysGlnAsnLysGluValTyrGlnGly 134
      ||| ::::: |||||
841 GCUAUCUUCUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 890
841 uValThrGluLeuSerProGlnLysLLeuLagLLeuLLeuSerThrGlyTyr 151
      ::::: |||||
891 CGUUCUGCGGAGAGUUAUACAGACGUGACCAAC..... 927
151 LysSerLLeuSerHisValLLeuLLeuSerLeuLysThrValLysGlyThr 167
      ::::: |||||
928 .....UUGAACCAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 969
168 LysGlnLeuLysLeuAspSerSerLLeuThrAspAlaLeuLLeuLysLys 184
      ::
970 GAA..... 972
184 sValAlaValGlyAspValLLeuTyrLLeuLLeuLLeuLLeuLLeuLLeu 198
      ||| ::::: |||||
973 .....GGGAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 1012
199 .....AlaValLysArgValGlyArgCysAspSer.....PheAla 210
      ||::: ||| |||||
1013 UUUUACUUCUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 1062

```

```

211 ThrGluTyrAspLeuGluValIleGluTyrValProIleProIleGlyGlu 227
    :|||:|||||:|||||:
1063 GGCGGCGCCGAGUUAAGGUGUGAAGCAACUUG.....AA 1100
    :|||:|||||:|||||:
227 uValHis...LysLysLysGluIleValGlnAspValThrLeuHisAspL 243
    :|||:|||||:|||||:
1101 AGUUCACGCGUAAGAACACCCUUAAGCAGAAAGUUGU.....GAU 1141
    :|||:|||||:|||||:
243 euAspAlaIleAsnIleGlnProGlnGly.....GlyGlnAspIleLeu 257
    :|||:|||||:|||||:
1142 UGAUAUUAUGUGGUCUACAAACCCAGGCGUUGUGUGUGUCUUAUUAGG 1191
    :|||:|||||:|||||:
258 SerLeuMetCylGln.....MetMetLysProArgLysThrGluIleTh 272
    :|||:|||||:|||||:
1192 AAUGUCUUAUGAAGCAGCUCUUAAGUGUGUGUGUGUGUGUGUGUGUG 1230
    :|||:|||||:|||||:
272 rGluLysLeuArgGlnGlnIleAsnLysValVal.....AsnArgTyrI 287
    :|||:|||||:|||||:
1231 .....AAUAAUCCGAUAUUAUGUUGUCUACAGAA 1258
    :|||:|||||:|||||:
287 leAspGluGlyIleAlaGluLeuValProGlyValLeuPheIleAspGlu 303
    :|||:|||||:|||||:
1259 UUAUUAAGACAGAAAGUAGAGUUAUUGUGUGAGCUCUUAAGAAAGUAG 1308
    :|||:|||||:|||||:
304 ...ValHisMetLeuAspIleGluCysPheSerTyrLeuAsnArgAlaLe 319
    :|||:|||||:|||||:
1309 ACAGUUCACAAAGAAAGACGAAUUGUGUGUCUAC...CAUGAGGACAG 1355
    :|||:|||||:|||||:
319 uGluSerProLeuSerProIleValIleLeuAlaThrAsnArgGlyIleC 336
    :|||:|||||:|||||:
1356 ACUACCAUUGUGUGU.....CUACUCUUGUGAGUCUGUGUGUGUGUG 1399
    :|||:|||||:|||||:
336 yAsnVal.....ArgGlyThrAspMetThrSerProHisGly 348
    :|||:|||||:|||||:
1400 AUAAGGUUAUUAUUGUACACGCGCGUGCAAGCGAAUACAUUAUGCA 1449
    :|||:|||||:|||||:
349 IlePro.....ValAspLeuLeuAspPar 356
    :|||:|||||:|||||:
1450 CUUCCUUAAGAGAGUCAAUUGCUUAUCAAAGAAUUAUUAAGAACAGA 1499
    :|||:|||||:|||||:
356 gLeu.....ValIleIleArgT 362
    :|||:|||||:|||||:
1500 AUUGGCGUGCUUAUUGGUGAGUGUGAGUCUAGCAAAUUAUUAUUAUG 1549
    :|||:|||||:|||||:
362 hArgLysThrTyrGly.....ProThrGluMetIle 371
    :|||:|||||:|||||:
1550 UCCAAACUACAGAGGCUUCAAACGACUUAAGACAGCAGCAAAUG... 1596
    :|||:|||||:|||||:
372 GlnIleLeuAlaIleArgAlaGlnValGluIleAspMetAspGluG 388
    :|||:|||||:|||||:
1597 .....GCAGCGUCCAAUGUGUACAGUACGUGUUAUGUGUAGAA 1634
    :|||:|||||:|||||:
388 uSerLeuAlaTyrLeuGlyIleGlyGlnGlnThrSerLeuArgHisA 405
    :|||:|||||:|||||:
1635 A.....CUUGGCCCGAGUA.....CAUUAUUAAGAGAAACCAUG 1666
    :|||:|||||:|||||:
405 la.....IleGlnLeuIleSerProAlaSerValValSerLysThrAsnGly 420
    :|||:|||||:|||||:
1667 CUUAUGCUUGUGACACAGUGCCUCAAUUAUUAUUAUUAUUAUUAUUA 1716
    :|||:|||||:|||||:
421 ArgGluLysIleCysLysAlaAspLeuGluGlnValSerGlyLeuTyr 437
    :|||:|||||:|||||:
1717 UAUCA.....AUUGAU...GAAGAGUUGUGUUAUUAUUAUUA 1751
    :|||:|||||:|||||:
437 uAspAlaLysSer...SerAlaArgLeuLeuGlnGlnGlnGluArgT 453
    :|||:|||||:|||||:
1752 UGAGGCGACGAAAUUAAGCUGCUUAUUAUUAUUAUUAUUAUUAUUA 1801
    :|||:|||||:|||||:
453 yr 453
    :|||:|||||:|||||:
1802 AC 1803
    :|||:|||||:|||||:

```

```

seq.name: /cgn2_6/protdata/2/ina/6B_COMB.seq:US-09-446-504-83
seq_documentation_block:
; Sequence 83, Application US/09446504
; Patent No. 6218150
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshimi
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazue
; APPLICANT: MIKAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshi
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/446,504
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 83
; LENGTH: 3574
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
US-09-446-504-83

alignment_scores:
    quality: 111.50      length: 367
    ratio: 0.664         gaps: 18
    percent similarity: 45.777    percent identity: 21.526

alignment_block:
US-09-589-510-4 x US-09-446-504-83 ..
Align seg 1/1 to: US-09-446-504-83 from: 1 to: 3574

54 AspmetiLeArgGlnLysLysMetAlaGlyArgAla..... 65
|||||:|||||:|||||:
2149 GACATTGTAGAGACAGACACATAGTGAAGGCTCAAGCACTACGTCAA 2198
66 .....ValLeuLeuAlaGlyProProAlaThrGlyL 76
|||||:|||||:|||||:
2199 AACTGATCAATGCCCACTACTCTTGCAGGCCCTGCTGCGAA 2248
76 yThrAlaLeuAlaLeuGlyIleAlaGlnGlnGlnGlySerLysValPro 92
|||||:|||||:|||||:
2249 AACTACAGCGCTTGGCCCTGCAAGACGCTT..... 2283
93 pheCysProMetValGlySer.....GluValTyr 102
|||||:|||||:|||||:
2284 .....TTCGCGGAAAACTGAGGCACTAATTCTCTGAGTGA 2321
102 rSerSerGluValLysLysThrGluValLeuMetGlu.....AsnP 116
|||||:|||||:|||||:
2322 TCCTTCAGTGAAGAGGTATTAACGTATTAGAGAAAGTTAAGCGT 2371
116 heArgArgAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGlu 132
|||||:|||||:|||||:
2372 TTGCGAGAACAAAGCCTATAGAGAGCAAGCTTCACATATTCTCTT 2421
133 GlyGluValThrGluLeuSerProGluGluAlaGluSer.....Th 146
|||||:|||||:|||||:
2422 GATGAGCGCGAGCTTAACTCAAGATGCCCAACAAAGCCTTAAGAAGAC 2471
146 rThrGlyGlyTyrAlaLysSerIleSerHisValIleIleSerLeuLysT 163

```

```

2472 CATGGAATGTTCTGAGTAACGTTGCTTATCTGAGCTGTAAGTACT 2521
169 hValIyGlyGlyThrIyGlnIeuIySLeuAspSerIleTyr..... 177
2522 CTTCCAGATTAATGACCCATACGCTAGATGTGCAATATTCGGCTTC 2571
178 .....AspAlaLeuIleIyGlnIySValAlaValGlyAspVal 190
2572 AGACCTCTCCCGGATGAGATATAGCCAGAGACTAAG..... 2610
190 ILeTyrIleGlnIleAsnSerGlyAlaValIySArgValIyArgCysA 207
2611 .....TACATGCGCGAAATGAGGCTTAGAG..... 2637
207 spSerPheAlaThrGluTyrAspIeuGlnIleGluTyrValProIle 223
2638 .....CTACTGAGAGAGTCTCCAGCAATACCTTAC.....ATA 2673
224 ProIyGlnIyGlnIleIySLeuIySLeuIyGlnIleAspValThrIle 240
2674 GCAGAGAGATATGAGAGAGACATATAACATCTGCAAGCTGCAGCAGC 2723
240 uHisAsp.....LeuAspAlaIleAsnA 248
2724 TCTAGACAAGAGATCACCAGCAAAACGATTCATGCTAGCAGTAGAG 2773
248 IagInProGlnIyGlyIleAsp.....IleLeuSerIleuMetGlyGln 262
2774 CTAGACCTGAGATATAGAGATGATGCTCTGCTCAAGGCGCAC 2823
263 MetMetIySProArgIySThrGluIleThrGlnIySLeuIyGlnIle 279
2824 TTCTTGAAGGCCAG.....GAAAGCTTAGGAGAGACT 2858
279 eAsnIySValIleAsnArgTyrIleAspGlnIyIleAlaIleuValP 296
2859 TCTCAAG.....CAAGGACTTAGT..... 2877
296 roGlyValIleuPheIleAspIleuIleIySLeuIleGlyCysPhe 312
2878 ..GGAGAGATGCTACTAGTCAGTCAGCAAA.....GAAGTCTC 2916
313 SerTyrIleuAsnArgAlaIleuGlnIleuSerProIleuValIlele 329
2917 .....AACCTGCCATAGAGAGCCAAAGAG.....GTTCTGCT 2951
329 uAlaThrAsnArgGlyIleCysAsnValArgGlyThrAspMetThrSerP 346
2952 TGCTGATTAAGATAGAGAGATTAACCTCAGACTCGTTGAA..... 2991
346 roHisGlyIleProValAspIleuAspArgIleuValIleIleArgThr 362
2991 ..... 2991
363 GluThrTyrGlyProThrGluMetIleGlnIleLeuAlaIleArgAla 379
2992 .....GGGCTAATGAATATATTCAGCTTAGAGACTCTTAGCACA 3032
379 n 379
3033 G 3033
seq_name: /cgn2_6/ptoddata/2/ina/5B_COMB.seq:US-08-993-228-11
seq_documentation_block:
; Sequence 11, Application US/08993228
; Patent No. 5976838
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallee, Edward R.
; APPLICANT: Racie, Lisa A.

```

```

APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael J.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,228
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-993-228-11

alignment_scores:
Quality: 108.50 Length: 542
Ratio: 0.448 Gaps: 25
Percent Similarity: 44.649 Percent Identity: 20.111

alignment_block:
US-09-589-510-4 x US-08-993-228-11 ..
Align seg 1/1 to: US-08-993-228-11 from: 1 to: 2837
23 GlyIeuGlyIleuAspAlaAsnGlyMetAlaIleAlaIleuAlaIagIyPh 39
|||||.....MetAl 62
671 GGCATTGGCCGGACAGGCCGAGGATGGGCGACTTTCAGT..... 712
39 eValGlyGlnAlaIleAla..... 45
|||||.....
713 .GTGCGAAGAACCACTGCCAAGCTTAAAGATGAATGATGTAAGT 761
46 ..ArgGlnAlaIleGlyIleuAla..... 52
|||||.....
762 TCAAGATGTGCTGCTGTGAGAGGCCCAAGCTAGAGATCATGATTT 811
53 ValAspMetIleArgGlnIySlys.....MetAl 62
|||||.....
812 GTGATTTCTTGAAAAACCAAGACAGTATCAAGACTAGAGACATTAAT 861
62 agIyArgAlaValIleuAlaIagIyProAlaThrGlyIySThrAlaIu 79
.....
862 CCAAGAGGTGCCATCTCTCAGCTGCTCCAGGCACTGGGAAACGCTGC 911
79 euAlaIeuGlyIleAlaGlnIleuGlySerIyValProPheCysPro 95
|||||.....
912 TACCTAAGGCCACAGCCGGAGAA.....GCCAATGTCCTTATCACC 955

```

```

96 MetValGlySerGluValTyrSerSerGluValLysLysThrGluValLe 112
   :: |||||
956 GTTAGTGATCTGAG.....TT 972
112 UmetGluAsnPharGargalaleGlyLeuArgLleLysGlu..... 126
   ::|||
973 TTGTGAGATGTCCTGTGTGGCCCTGCTAGAGTCGAGACTTATTTGG 1022
127 .....AsnLysGlu.....ValTyrGluGlyLysValThr 136
   ::|||
1023 CCTTGCTCGGAGAGATGCCCTTGATCTCTTCATCGATGAATGAT 1072
137 GluLeuSerProGluGluValGluSerThrThrGlyLysValLys.. 152
   ::|||
1073 GCGGTGGAGAGAGAGAGAGAGAGACCTTTGGACGAGAGATGCA 1122
153 .....Ser 154
1123 GGAGACACACTCAACCACTGCTGTGGAGATGATGTTTAAATCA 1172
154 leserHisValLleleSerLeuLysThr..... 163
   ::|||
1173 CAACAATGTCATTTTGGCCGACCAATGACAGACAGACCAACGAC 1222
164 ValLysGly.....ThrLysGluLeuLysLys 172
   ::|||
1223 ATAAAGAGAGAGCTTCTATTTTCAAGTTTCATGCTCGACCGCTAA 1272
172 uAspSerSerLetyrAspAlaLeuLleLysGluLysValAlaVal... 187
   ::|||
1273 GGACAGTACCTCGAGAGAGATAAATTTGGCAAGAACTGCGATCTTAA 1322
188 .....GlyAspValLetyrLleGluLleAsnSerGly 198
   ::|||
1323 CTCAGGGTTTTCAGGTGCTGATGTTCTATGCTGTATGAAGCTGCG 1372
199 AlaValLysArgValGlyArgCysAspSerPheAlaThrGluTyrAsp 215
   ::|||
1373 TTGATGCTCGAAGCACTGCTGATTCATTAATCAGAAACCTTTGA 1422
215 uGluAlaGluGluTyrValProLleProLysGlyGluValHisLysLys 232
   ::|||
1423 ACAGGCAATGAGCGAGTG.....ATTGGGCTTAAAGAAAAA 1463
232 ysgLileValGlu.....AspValThrLeuHisAspLeu 243
   ::|||
1464 CGCAGGTTTCACCTGAGAGAGAAAGACTGTGATACCCAGAGCA 1513
244 AspAlaAlaAsnAlaGlu..... 249
1514 GGCCATGCGGTGGCGGCTGATCTGAGCAGCAGACGCCCTTTTAA 1563
250 .....ProGlnGly..... 252
1564 GGTATCATCATCCACGTCGCAAGAGACTAGTTATGCTACATTTTAC 1613
253 .....GlyGluAspLleLeuSerLeuMetGly 261
1614 CAAAGAGACATACCTTATACCAAGAGCAGCTTTGGATAGAGATGTG 1663
262 GlnMetLysProArgLysThrGlu.....Leth 272
   ::|||
1664 ATGACTTTAGGTGTCAGTCTGAGAGAAATCTTTTGGAGAAATTAC 1713
272 rGluLysLeuArgGluGluLleAsnLysValValAsn.....A 285
1714 AACTGTGCTCAAGATGACTTGAGAAATCACTCAGAGCTATATGCC 1763
285 rGTYrLleAspGluGlyLleAlaGluLeuValProGlyValLeuPhe 301
1764 AAATGTTTCAGTTTGGCATGAATGAAGGTT...GGCAAAATCTCTT 1810
302 AspGluValHisMetLeuAspLleGluCysPheSerTyrLeuAsnArg 318

```

```

1811 GACCTCCACAGTCAGGGGAGCATG.....GT 1836
   ::|||
318 aleuGluSerProLeuSerProLleValLleLeuAlaThrAsnArgGly 335
   ::|||
1837 ATTGGAGAAACCTTACAGTGA.....GCCACTGCAGATTTGA 1874
335 leCysAsnValArgGlyThrAspMetThrSerProHisGlyLleProVal 351
   ::|||
1875 TAGATGATGACATACCAATACTTATTAAATGATGCTTATTAAGAAC 1924
352 AspLeuLeuAspArg.....LeuValLle 359
1925 GCTCTTCTCAGAGAAAGAAAGCTGACGTGAGAGAGTTGCTCTGTG 1974
359 eLleArgThrGluThrTyrGlyProThrGluMetLleGluLleValLle 376
   ::|||
1975 GTTGAAGAAAGATTTAGATGAATGATATGATTTGAATTTTGGGCC 2024
376 leArgAlaGluValGluGluLleAspMetAspGluLeuValLys 392
   ::|||
2025 CCAGACCATTTGGGAGAAA...TCTACTATGAAGATTTGTGAGAGC 2071
393 leuGlyGluLleGlyGlnThrSerLeuArgHisAlaLleGluLleu 409
   ::|||
2072 ACTGGCAGCTGTGATGAGACACTCATT..... 2101
409 eSerProAlaSerValLysLysThrAsnGlyArgGluLysLleCysL 426
   ::|||
2102 ...CCAGAGGCTTAAAGACTGAGACAGAGGCGGAGAAAG..... 2140
426 yAlaAspLeuGluGluValSerGly 434
2141 ....GAGAAAGAGAGACCCCGGGT 2161

```

seq_name: /cgn2_6/plodata/2/ina/6B_COMB.seq:US-09-103-840A-2

```

seq_documentation_block:
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTNER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

alignment_scores:

Quality:	Ratio:	Length:
133.50	0.746	351
Percent Similarity:	50.997	Gaps: 18
		Percent Identity: 22.222

alignment_block:

US-09-589-510-4 x US-09-103-840A-2 ..

Align seg 1/1 to: US-09-103-840A-2 from: 1 to: 4403765

```

36 AlaIaagIlyPheValIGlyGlnAlaAlaIaargIuValaIaagIlyLeuAl 52
||| |||.....:~::~||| |||
1669711TCCGAGGAGGCTGGCCGAGGTGCACACCGCTGGAGCGGCATCTTCCGA 1669560
52 aVal.....AspMetIleargGlnLysLysMetA 62
||| |||.....:~::~||| |||
1669561GGTCACAGCGGATTCGTGCGCCGACGACAGCTGTGTGAGCGGATGCTCG 1669610
62 IaGIaIaIaIa.....ValLeuLeuAlaGIyProProAlaIaThr 74
||| |||.....:~::~||| |||
1669611TCCGCGCTGCTCCAGAGCGGCGATGCTGCTGAGGCGCTTCCGCGGCTG 1669660
75 GlyIysThrIaLeuAlaLeuAlaGlyIleAlaGlnGluLeuGly..... 88
||| |||.....:~::~||| |||
1669661GGTCAGAGCTTGGCGGTGAGACCTTGTGCTGGGTGCTGCGCGGACATT 1669710
89 .SerIysValIProPheCysProMetValIGlySerGluValIYrSerSerg 105
||| |||.....:~::~||| |||
1669711TCCGCGCATCCAGTTCACCCCG.....G 1669733
105 LuValIYsIysThrGluValIleuMetIuAsnPhaIaIaIaIaIaIaGly 121
||| |||.....:~::~||| |||
1669734ATCTGTCGCCACCGCATC.....ATCGGG 1669759
122 LeuArgIle..LysGluAsnLysGluValIYrGluGlyGluValIThrGI 137
||| |||.....:~::~||| |||
1669760AGCGCATCTACCGGCAAGGAGGAGCAAAATCCAC.....ACCGA 1669800
137 uLeuSerProGluGluAlaGluSerThrThrGIyGIyAlaLysSerI 154
||| |||.....:~::~||| |||
1669801ACTCGAGACCGGTGTGGCCAACTCTCG.....CTCGCGACAGAGA 1669841
154 LeSerHisValIleIleSerLeuLysThrValIYsGIyThrLysGlnLeu 170
||| |||.....:~::~||| |||
1669842TTCACCGCGGCT.....CCGCGC 1669858
171 LysLeuAspSerSerIleYrAspAlaIaIeIleYsGlnLysValaIaVa 187
||| |||.....:~::~||| |||
1669859AAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1669908
187 IGIAspValIleYrIleGluAlaAsnSergIyAlaValIYsArgValG 204
||| |||.....:~::~||| |||
1669909CGCGGTAGACCTTCCGATGCCACCGCTGCTGCTG..... 1669948
204 IYArgCysAspSerPheAlaThrGluYrAspLeuGluAlaGluGluYr 220
||| |||.....:~::~||| |||
1669949.....ATGGCGACGCAAGAACCGGATCGACGACGAGCGGCTC 1669984
221 ValProIleProLysGlyGluValIHisLysLysLysGluIleValGlnAs 237
||| |||.....:~::~||| |||
1669985TACCGCGTACCGGAGCGCAA.....CGGCA 1670010
237 pValIThrLeuHisAspLeuAspAlaAlaAsnAlaGlnProGlnGlyG 254
||| |||.....:~::~||| |||
1670011CCGCTTCTTCAAGATCAACGTGGCTACCGCTGCGCCGAGAGAGAGC 1670060
254 InaSPIleLeuSerLeuMetGlyIleMetLysProAlaGlySfThGIu 270
||| |||.....:~::~||| |||
1670061CGCAATCATCTACCGTATGGGT...GTACCGCGCGCGAGCGCAACAG 1670107
271 IleThr.....GluLysLeuArg...GlnGluIleAsnLysVala 283
||| |||.....:~::~||| |||
1670108ATCTCGACACGCGGCGACCTGCTGCGGTGAGAGATA.....GGGCG 1670151
1670152CAACAACCTTCTCACACACGCGGTGTGACTATGTCTGACTGCTCT 1670201
300 heIle.....AspGluValIHisMetLeuAspIleGluCysPhe 312
||| |||.....:~::~||| |||
1670202TTCGCCACCCGCAAAACCGAGAGTGGGAGTGAAGAGAGAGCTGG 1670251
313 SerTyrlleAsnArgAlaIaLeuGluSerProLeuSerProIleValIle 329

```

```

1670252GTTCGCTTGGC.....GCATCCCGCGCGTTCGCTGCGGCATCAT 1670292
329 uAlaThrAsnArgGlyIleCysAsnValArgGIyThrAspMetIhrSerP 346
||| |||.....:~::~||| |||
1670293CGCGCGGCGGTCCTGCGCTGCTGCGGCGGTGACTATGTTCATCC 1670342
346 roHisGIy.....IleProValAspLeuLeuAspArgLeuVal 358
||| |||.....:~::~||| |||
1670343CGCAACCTCATTCAGGATTCATCTGATGCTGTGGACACCGGCTCGTG 1670392
359 Ile 359
1670393CTC 1670395
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-074-579-2
seq_documentation_block:
: Sequence 2, Application US/09074579
: Patent No. 6001596
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Guejler, Karl J.
: APPLICANT: Patterson, Chandra
: TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
: TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/074,579
: FILING DATE: HEREMITH
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Ceirone, Michael C
: REGISTRATION NUMBER: 39,132
: REFERENCE/DOCKET NUMBER: PF-0505 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3636 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: UTRNOT02
: CLONE: 688183
US-09-074-579-2
alignment_scores:
Quality: 104.50 Length: 484
Ratio: 0.525 Gaps: 23
Percent Similarity: 41.116 Percent Identity: 17.975
alignment_block:
US-09-589-510-4 x US-09-074-579-2 ..

```

Align seg 1/1 to: US-09-074-579-2 from: 1 to: 3636

```

93 PheCysProMetValIglySerGluValTyrSerSerGluValLysThr 109
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
380 TTCACCTATGCTTATTGGAGACAGAGTGATACGGCCAAATTAACAGAG 429
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 rGluValLeuMetCLeuSphenArgAlaIleGlyLeuArgIleLysG 126
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430 AGAA.....AGAAAGAGTGATAGGGTAAG 458
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
126 LuAsnLys.....GluVal 130
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
459 AGAAAGAGATAAACCACAGACAAATGAGAGAAAGGCGACTGAATA 508
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 TyrGluGlyGluValThrGluLeuSerProGluGluAla..... 143
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
509 TTCAGAGCTCTGCAGTATCCACAGACAGACAAACCCCTTTTCT 538
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
144 .....GluSerThrThrGlyGlyTyrAlaLysSerI 154
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
559 GAGTTATGAGAGCTCTGCAGAGGCGCTGGCAAGTACGACACAGC 607
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 IeSerHisValIleIeSerLeuLysThrValLysGlyThrLysGluLeu 170
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
608 .....ATCAGCGTGCGGCCCCAGCAGCTCTCCGGAGGCTG 643
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 LysLeuAspSerSerIleTyrAspAla..... 179
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
644 AGCGTGACCGTAATATCTCGAGAGCGCGGCATCCATCCCTGGAGGT 693
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 ..... 179
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
694 GCTGCCCTTCACACAGCAGAGGGGCACTGGCGCGGGAGAGATG 743
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 .....LeuIleLysGluLysValAlaIleValGly 188
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
744 ATTCTGGGCTCCGCCCTACTGCTATTAACCAAAATGAACATTGGCC 793
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 AspValIleTyr.....IleGluAlaAsnSerGlyAlaValLysArg 203
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
794 AACATATTTTAACTACTAGTACACAAACAGCAGATTCCCGAGAA 843
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 IglyArgCysAspSerPheAlaThrGluTyrAspLeuGluAlaGluLys 220
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
844 TGAATTTTGGAGACTTATCATATGATATGATGATAGAGAGACAGA 893
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 yVal..... 221
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
894 GCATGGGAGATCCAGGTTCTAATGCTATTGTGCACTACTTGTCT 943
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 .....ProIleProLysGly..... 226
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
944 CCAAGAGACTTCTCTCTTACCACCAAGATGTGATTTGCTGTGACAG 993
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
227 .....GluValHisLysLysGluIleValG 236
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
994 CAGTCTCTATGATGGAGAACCAACTCCGGCAGACCAAGATGCCCTCT 1043
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 IAspValThrLeuHisAspLeuAspAlaAlaAsnIleGluProGluGly 252
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1044 TCACAAT...CTCCATGACCTC.....CGACCC..... 1069
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 GlyGluAspIleLeuSerLeuMetGly..... 261
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1070 ...CAGGACCGTTTCAGTATCATGATTTTCCAAACCGGATCAAGATAG 1116
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 ..... 261
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1117 GAAGACCACTTGATATCATCATCCACAGACATCAGGAGTGGGAAG 1166
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
262 .....GlnMetMetLysProAlaGlyThrGluIleThrGluLys 274
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1167 TGTACATTCACATATATGTCAACCCACTGAGGACACAGACATCAAGGGGCC 1216
1217 CTCACAGAGGCCATC...AGGCTCCCAACAGATGATGCGCCACAGATG 1263
290 yIleAlaGlu.....LeuValProGlyValL 299
1264 CATTTGAGACCGGAGCGGTCCCTCATGCTTCTTCCAGAGATGGGAAGC 1313
299 euPheIleAspGluValHisMetLeuAspIleGluCysPheSerTyrLeu 315
1314 CCAGGCTCGGGAGAGCGCACACCTCAAGATC.....CTC 1348
316 AsnArgAlaLeuGluSerProLeuSerProIleValIleLeuAlaThrAs 332
1349 AACAAACCCCGAGAGGCCCGCCGAGGCAAGTCTCATCTTACCATTTGG 1398
332 nArgGlyIleCysAsnValArgGlyThrAspMetThrSerProHisGlyI 349
1399 CATGGC.....AACGACG 1412
349 IeProValAspLeuLeuAspArgLeuValIleIleArgThrGluThrTyr 365
1413 TGACTTCAGGCTGCTGGAGAACTGTGCTG.....GAGAACTGT 1453
366 GlyProThrGluMetIleGluIleLeuAlaIleArgAlaGluValGlu 382
1454 GGCTTCACACGGCGGCTGCAC.....GAGGA 1479
382 uIleAspMetAspGluGluSerLeuAlaTyrLeuGlyLysIleGlyGln 399
1480 GGAGAGCGAGCGCTCGACGCTCATCGGGTCTACAGATGAATC.....A 1523
399 InThrSerLeuArgHisAlaIleGluLeu...IleSerProAlaSerVal 414
1524 GGACCCCGCTCCTCTGTGACATCGCATGATTAATCCCGACGCTCACTG 1573
415 ValSerLysThr.....AsnGlyArgGly 423
1574 GTGAGGCCACCAAGACCTGTTCCCACTACTTCAACGGCTGGAGAT 1623
423 sIleCysLysAlaAspLeuGluGluValSerGlyLeuTyrLeuAspAla 440
1624 CATC.....ATTGCGGGAGAGCTGGTGACACAGA 1652
440 ys 440
1653 AG 1654

seq_name: /cgn2_6/ptodata/2/ina/6b_COMB.seq:US-09-388-774-2
seq_documentation_block:
; Sequence 2, Application US/09388774
; Patent No. 6228991
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; INHIBITOR HEAVY CHAIN PRECURSOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```


159 GCGCCCTTCACTTCCTCCCGGGCTTTTCCTCCGACATTACCACTTTAGG 208
159 GCGCCCTTCACTTCCTCCCGGGCTTTTCCTCCGACATTACCACTTTAGG 208
32 aAlaAlaLeuAlaAlaGlyPheValGlyIn.....AlaAlaAla 46
32 aAlaAlaLeuAlaAlaGlyPheValGlyIn.....AlaAlaAla 46
209 GCGCTCTCGAGAGTCGGAGACCCCGCGGAAGACGGGGGCGCGGTCC 258
209 GCGCTCTCGAGAGTCGGAGACCCCGCGGAAGACGGGGGCGCGGTCC 258
46 rgluAlaAlaaglyLeuAlaVal..... 53
46 rgluAlaAlaaglyLeuAlaVal..... 53
259 GCTCCGGCGCTCGGATGGGAAGTGTGTGGAGACCGACCCGGATGTTTCAG 308
259 GCTCCGGCGCTCGGATGGGAAGTGTGTGGAGACCGACCCGGATGTTTCAG 308
54AspMetLeaArgGlnLysMetAlaGlyArgAlaVal... 66
54AspMetLeaArgGlnLysMetAlaGlyArgAlaVal... 66
309 TGTGATGCCACAGTTGCTGGCGCTGTTCAAGCGGTGGCGGAGCTGTCAg 358
309 TGTGATGCCACAGTTGCTGGCGCTGTTCAAGCGGTGGCGGAGCTGTCAg 358
67LeuLeuAlaGlyProProAlaThrGlyLysThAlaAl 79
67LeuLeuAlaGlyProProAlaThrGlyLysThAlaAl 79
359 AAAGGTACTCTTTTGATGTGGTGGAGCTTGATATGCTGTGTAAACCGCAA 408
359 AAAGGTACTCTTTTGATGTGGTGGAGCTTGATATGCTGTGTAAACCGCAA 408
79 eAlaAlaGlyLLeaAlaGlnGlyLeuGlySerLysValProPheCysPro 95
79 eAlaAlaGlyLLeaAlaGlnGlyLeuGlySerLysValProPheCysPro 95
409 CAGCAAGAGGATTCACAGAGAAATNCCCGAAGATCA.....GCTCCT 452
409 CAGCAAGAGGATTCACAGAGAAATNCCCGAAGATCA.....GCTCCT 452
96 MetValGlySerGlyValLysSerSerGlyValLysLysLysThGluValLe 112
96 MetValGlySerGlyValLysSerSerGlyValLysLysLysThGluValLe 112
453 ACTGTGGATTTTTCAAAATTAACCTTAGACACAGAAAGTTTGAAGTCCAC 502
453 ACTGTGGATTTTTCAAAATTAACCTTAGACACAGAAAGTTTGAAGTCCAC 502
112 uMetGluAsnPheArgAlaAlaGlyLeuArgGlyLeuGluAsnLysG 129
112 uMetGluAsnPheArgAlaAlaGlyLeuArgGlyLeuGluAsnLysG 129
503 CATC.....TTTGACTTGGAGAGTGGATTAAGAATTGCGGGAAATCTGGA 546
503 CATC.....TTTGACTTGGAGAGTGGATTAAGAATTGCGGGAAATCTGGA 546
129 luValTyrGluGlyGlu.....ValThrGluLeuSer 139
129 luValTyrGluGlyGlu.....ValThrGluLeuSer 139
547 AGAATTACTATGACTGCAATCAATCATGCGGTAAATTTGTTGTGATTCACgT 596
547 AGAATTACTATGACTGCAATCAATCATGCGGTAAATTTGTTGTGATTCACgT 596
140 ProGluGlu...AlaGluSerThrThrGlyGlyTyrAlaLysSerLLeSe 155
140 ProGluGlu...AlaGluSerThrThrGlyGlyTyrAlaLysSerLLeSe 155
557 GATGAAGACAAATTCGAAGACACAAACAGAGCTATGTACGAATTCCTAG 646
557 GATGAAGACAAATTCGAAGACACAAACAGAGCTATGTACGAATTCCTAG 646
155 rHisValLeuLeuSerLeuLys.....ThrValLysGlyThrLysGlnL 170
155 rHisValLeuLeuSerLeuLys.....ThrValLysGlyThrLysGlnL 170
747 ACAATTCCTAGATATTCGGGAAAGCCTATATGTGTGGCAATATACAG 696
747 ACAATTCCTAGATATTCGGGAAAGCCTATATGTGTGGCAATATACAG 696
170 eLys.....LeuAspSerSerLLeTyr 177
170 eLys.....LeuAspSerSerLLeTyr 177
697 ATTAAGAGAGGAGCTTACGAGAGAGCTGATGTCATGATGCTATCTCTG 746
697 ATTAAGAGAGGAGCTTACGAGAGAGCTGATGTCATGATGCTATCTCTG 746
178 AspAlaLeuLeuLysGlnLysValAlaAlaGlyAspValLLeTyrLLeG 194
178 AspAlaLeuLeuLysGlnLysValAlaAlaGlyAspValLLeTyrLLeG 194
747 GAATAATTTGGTCATGTAGCAACAATGCTGTGTGTCAGATAGAACCATGTTC 796
747 GAATAATTTGGTCATGTAGCAACAATGCTGTGTGTCAGATAGAACCATGTTC 796
194 uAlaAsnSerGlyAlaValLysArgValGlyArgCysAspSerPheAlaT 211
194 uAlaAsnSerGlyAlaValLysArgValGlyArgCysAspSerPheAlaT 211
797 AGCATATCGGGGTATGGAAAGAAATTTGACAG..... 830
797 AGCATATCGGGGTATGGAAAGAAATTTGACAG..... 830
211 hrGluTyrAspLeuGlnAlaGluGluTyrValProLLeProLysGlyGlu 227
211 hrGluTyrAspLeuGlnAlaGluGluTyrValProLLeProLysGlyGlu 227
831TTC 833
831TTC 833
228 ValHisLysLysLysGluLLeValGlnAspValThrLeuHisAspLeuAs 244
228 ValHisLysLysLysGluLLeValGlnAspValThrLeuHisAspLeuAs 244
834 ATTAATAAAGAGCTTATTTGCTGCTCATGATTAATTGCAGAGACTTTGA 883
834 ATTAATAAAGAGCTTATTTGCTGCTCATGATTAATTGCAGAGACTTTGA 883
244 rAlaAlaAsnAlaGlnProGlnGlyGlnAspLLeLeuSerLeuMetC 261
244 rAlaAlaAsnAlaGlnProGlnGlyGlnAspLLeLeuSerLeuMetC 261
884 TGCCTTAATTAAGACGCTCCAAATAAGACACACAGACAGCAGCTGCTTCG 933
884 TGCCTTAATTAAGACGCTCCAAATAAGACACACAGACAGCAGCTGCTTCG 933
261 LysImeLysLysProArgLysThrGluLLeThrGluLysLLeArgGln 277
261 LysImeLysLysProArgLysThrGluLLeThrGluLysLLeArgGln 277
934 AGGAACAAGAGAAACAAGAAAGAGCTGAACAGATTCGGAATAATTCAGGAA 983
934 AGGAACAAGAGAAACAAGAAAGAGCTGAACAGATTCGGAATAATTCAGGAA 983
278 Glu.....LLeAsnLysValValAsnArgTyrLLeAsp.....GluG 290
278 Glu.....LLeAsnLysValValAsnArgTyrLLeAsp.....GluG 290

984 GAAAGAAAACAAATGTGACAGAGCAAGGCTGAACCTCGATGGAAACCAAGTGG 1033
290 YLEAlaGluLeuValPro 296
|::| | | | | | | | | |
1034 TCTGGCTGAGTTGGACCA 1052

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 06:56:39 ; Search time 25.58 seconds

(without alignments)
2353.354 Million cell updates/sec

Title: US-09-589-510-4

Perfect score: 2263
Sequence: 1 MRIEVQSTKKQRIATHH.....YLDKSSARLLQEQEERYIT 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: SPREMBL_16:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.unclassified:*
14: sp.vertebrate:*
15: sp.virus:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2039	90.1	458	10 Q9M326	Q9m326 cicer ariet
2	1994	88.1	458	10 Q9PMB9	Q9Pmb9 arabidopsis
3	1723.5	76.2	456	4 Q9Y265	Q9Y265 homo sapien
4	1722.5	76.1	456	11 Q35753	Q35753 rattus norv
5	1719.5	76.0	456	13 Q9DE26	Q9DE26 xenopus lae
6	1655.5	73.2	456	5 Q9VH07	Q9Vh07 drosophila
7	1643.5	72.6	456	5 Q9NH53	Q9Nh53 drosophila
8	1517.5	67.1	463	3 Q03940	Q03940 saccharomyc
9	1311	57.9	458	5 Q17607	Q17607 caenorhabdi
10	989.5	43.7	465	3 Q94692	Q94692 schizosacch
11	989	43.7	462	13 Q9DE27	Q9DE27 xenopus lae
12	989	43.7	481	5 Q9V3K3	Q9V3K3 drosophila
13	986.5	43.6	441	1 Q9V1R4	Q9V1R4 pyrococcus
14	975	43.1	441	1 Q59468	Q59468 pyrococcus
15	974	43.0	463	11 Q9WTM5	Q9Wtm5 mus musculu
16	973	43.0	463	4 Q9Y230	Q9Y230 homo sapien
17	963.5	42.6	469	10 Q9FCW0	Q9Fcw0 arabidopsis
18	942.5	41.6	449	1 Q28462	Q28462 archaeoglob
19	932.5	41.2	471	3 Q12464	Q12464 saccharomyc

20	928	41.0	473	10 Q9M2X5	Q9m2x5 arabidopsis
21	895	39.5	442	4 Q9Y361	Q9Y361 homo sapien
22	888.5	39.3	473	1 Q9YFB5	Q9Yfb5 aetopyrum p
23	863	38.1	448	5 Q9G2H2	Q9g2h2 caenorhabdi
24	417	18.4	118	5 Q9GK8	Q9gk8 leishmania
25	299.5	13.2	112	1 P96020	P96020 sulfolobus
26	180.5	8.0	795	1 Q9V1N2	Q9V1n2 pyrococcus
27	179.5	7.9	798	1 Q59515	Q59515 pyrococcus
28	177	7.8	840	1 Q58419	Q58419 pyrococcus
29	171	7.6	835	1 Q9Y2X2	Q9Y2x2 pyrococcus
30	168	7.4	840	1 Q9Y2Z7	Q9Y2z7 pyrococcus
31	166.5	7.4	733	1 Q28972	Q28972 archaeoglob
32	159	7.0	1251	10 Q9MA34	Q9ma34 arabidopsis
33	153.5	6.8	699	1 Q9Y910	Q9Y910 aetopyrum p
34	148.5	6.5	726	1 Q9YC86	Q9Yc86 aetopyrum p
35	146.5	6.5	811	1 Q28182	Q28182 archaeoglob
36	142.5	6.3	614	8 Q19922	Q19922 cyanidium c
37	141.5	6.3	892	11 Q9Z2K7	Q9z2k7 mus musculu
38	141	6.2	620	4 Q9H8W7	Q9h8w7 homo sapien
39	140	6.2	737	1 Q9HHW2	Q9hhw2 halobacteri
40	139.5	6.2	732	1 Q27676	Q27676 methanobact
41	138	6.1	809	3 Q60058	Q60058 schizosacch
42	137.5	6.1	709	2 Q9JTS1	Q9jts1 neisseria m
43	136.5	6.0	943	5 P91638	P91638 drosophila
44	136.5	6.0	944	5 Q9VS62	Q9vs62 drosophila
45	136	6.0	759	1 Q9HPU1	Q9hpu1 halobacteri

ALIGNMENTS

RESULT 1

Q9M326 PRELIMINARY: PRT; 458 AA.

```

AC Q9M326;
ID Q9M326;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE PUTATIVE RUV DNA-HELICASE.
OS Cicer arietinum (chickpea). (Garbanzo).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicot; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Cicer.
OX NCBI_TaxID=3827;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=CV. CASTELLANA; TISSUE=ETIOLATED EPICOTYLS;
RA Dopico B., Esteban R., Labrador E.;
RT "A RuvB DNA-helicase like protein is expressed in chickpea
epicotyls.";
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ276264; CAB76908.1; -.
DR InterPro; IPR001984; -.
DR InterPro; IPR003593; -.
DR PRINTS; PR00830; ENDOLAPTASE.
DR SMART; SM00382; AAA; 1.
DR Helicase.
SQ SEQUENCE 458 AA; 50262 MW; 61EAD73060376764 CRC64;

```

Query Match 90.1%; Score 2039; DB 10; Length 458;

Best Local Similarity 87.5%; Pred No. 1.1e-109; Indels 0; Gaps 0;

Matches 398; Conservative 33; Mismatches 24; Indels 0; Gaps 0;

```

QY 1 MRIEVQSTKKQRIATHHINKIGLDANGMALAAGFVGQAAAREAGLAVIDMIRKK 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 4 MKIEVQSTKKQRIATHHINKIGLEVGSKALPFASGFVGQAEAREACGLVDMIRKK 63
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 MAGRAVLLAAPPATGKTALALAGIAOLGSKVPPCPVWGSEYVSSVKKTEVLMENRRRI 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 64 MACKALLLGGPPGTGTALALGICQLGTFKPPVWGSEYVSTEYKTEVLMENRRRI 123
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 GLIRIKNKVEYGEVTELSPEAEASTTGGYAKSISHTVITSLTKVGTQKLKIDDSIYDAL 180

```

Db	124	GLRIKENKEVEGEVETELSPETESLNGVYKSTISHVILGKITVKGKOLKLDPTIYDAL	183
QY	181	IKKAVAGDVIIYIANSQAVKRVGRCDSEFATEYULDEAEVVPILPKGEVHKKEIYODVTL	240
Db	184	IKKAVAGDVIIYIANSQAVKRVGRSDAFAPFEPLDEAEVVPILPKGEVHKKEIYODVTL	243
QY	241	HDLDANNOPOGGODIILSLMGOMKKPRKTEITTEKLROINKVVRUYIDEBIAEIVPQVLF	300
Db	244	HDLDANNRPOGGODIILSLMGOMKKPRKTEITTEKLROINKVVRUYIDESVAELVPPVLF	303
QY	301	IDEVMLDIDIECFSYLNRALESPLSPIYILANRIGICVNRGTDMTSPHIGIVDLDRLVII	360
Db	304	IDEVMLDIDIECFSYLNRALESPLSPIYIFANRIGICVNRGDMTSPHIGIVDLDRLVII	363
QY	361	RTETYGPIEMTIOILAIRAQVEETIMDEBSLAYLGEIGQOTSLRHAIDLISPAVSVKYNG	420
Db	364	RTOETGPAEIIIOILAIRAQVEELVJADEBSLAFGEIGORTSLRHAIVDLSPASVAKING	423
QY	421	REKICKADLEVSGLYIDAKSSARULLDEOQRYIT	455
Db	424	RDNICKADLEETCSYIDAKSSAKULLDEOQRYIS	458

	RESULT	2	
09FM9	09FM9	PRELIMINARY;	PRN; 458 AA.
AC	09FM9;		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	ROY DNA-HELICASE-LIKE PROTEIN.		
OS	Arabidopsis thaliana (Mouse-ear cross).		
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
OC	Magnoliophyta; eudicotyledons: core eudicots: Rosidae; eurosids II;		
OC	Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-COLUMBIA;		
RX	MEDLINE-98162728; PubMed-9501997;		
RA	Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,		
RA	Tabata S.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. III.		
RT	Sequence features of the regions of 1,191,918 bp covered by seventeen		
RT	physically assigned P1 clones."		
RL	DNA Res. 4:401-414(1997).		
DR	EMBL; AB007651; BAB08331.1; -.		
DR	Helicase.		
SO	SEQUENCE.	458 AA;	50323 MW; 7DA33A99AFA365F CRC64;

Query Match	88.1%;	Score 1994;	DB 10;	Length 458;
Best Local Similarity	85.1%;	Pred. No. 4.4e-107;		
Matches 387;	Conservative 39;	Mismatches 29;	Indels 0;	Gaps 0;

[illegible]

Db	244	QDLDANANRPOGGDDIILSMQMKPKRTETITDKLRQELNKVARYIIDEQVAELYPGLF	303
Qy	301	IDEVMLDIECFSYLNRALESPLSYVLATNRGICVNRGDMTSPHGIIPVDLRLVII	360
Db	304	IDEVMLDMCEPSTYLNRALESLSPIYIFATNRGVCNVRGDMSPHGPVPTDLLDRYII	363
Qy	361	RTEYGPTEMIQIIAIRAQVEIEMDEESLAVLGEICQOTSLRAHILISPAVSXKNG	420
Db	364	RQIYDPESEMIQIIAIRAQVEELVYDEDECLVLGEIQRISLRAHVLLSPASIVAKNG	423
Qy	421	REKICKADLEEVSGLYLDPAKSSARLLQEOQERYIT	455
Db	424	KDNICKADIEEYTSIYLDPAKSSAALLMHQEQKYS	458

RESULT	3			
ID	09Y265	PRELIMINARY:	PRT:	456 AA.
AC	09Y265;			
DT	01-NOV-1999 (TRMBLrel. 12, Created)			
DT	01-NOV-1999 (TRMBLrel. 12, Last sequence update)			
DT	01-MAR-2001 (TRMBLrel. 16, Last annotation update)			
DE	ERYTHROCYTE CYTOSOLIC PROTEIN OF 54 KDA, ECP-54.			
CN	TP149.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BONE MARROW;			
RX	MEDLINE=98447618; PubMed=9774387;			
RA	Kudlick M., Salzer M., Prohaska R.;			
RT	"An eukaryotic RuvB-like protein (RUVBL1) essential for growth.";			
RL	submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98447618; PubMed=9774387;			
RA	Qiu X.B., Lin Y.L., Thome K.C., Plan P., Schlegel B.P., Weremowicz S.,			
RA	Parvin J.D., Dutta A.;			
RT	"An eukaryotic RuvB-like protein (RUVBL1) essential for growth.";			
RL	J. Biol. Chem. 273:27786-27793(1998).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLORECTAL ADENOCARCINOMA;			
RX	MEDLINE=99061945; PubMed=9843967;			
RA	Bauer A., Huber O., Kemler R.;			
RT	"Pontin5, an interaction partner of beta-catenin, binds to the TATA			
RT	box binding protein.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:14787-14792(1998).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER;			
RX	MEDLINE=98249797; PubMed=9588198;			
RA	Makino Y., Mimori T., Kolke C., Kanemaki M., Kurokawa Y., Inoue S.,			
RA	Kishimoto T., Tamura T.;			
RT	"Rip49, homologous to the bacterial DNA helicase RuvB, acts as an			
RT	autoantigen in human.";			
RL	Biochem. Biophys. Res. Commun. 245:819-823(1998).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=PANCREAS;			
RX	MEDLINE=99032802; PubMed=9813143;			
RA	Holzmann K., Gerner C., Korosec T., Poeltl A., Grimm R., Sauermann G.,;			
RT	"Identification and characterization of the ubiquitously occurring			
RL	nuclear matrix protein NMP 238.";			
RT	Biochem. Biophys. Res. Commun. 252:39-45(1998).			
DR	EMBL: Y18418; CAB46271.1; -			
DR	EMBL: AF070735; AAC77819.1; -			
DR	EMBL: AF090904; AAD04427.1; -			
DR	EMBL: AB012122; BAA28169.1; -			
DR	EMBL: AJ010058; CAA08986.1; -			

09NH53
ID 09NH53 PRELIMINARY; PRT; 456 AA.
AC 09NH53;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE PONTIN.
OS PONTIN OR CG4003.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phryganea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN NCB1_Taxid=7227;
[1]
RP SEQUENCE FROM N.A.
RA Chauvet S., Useglio F., Missod R., Aragnol D., Pradel J.;
RT "Control of Mt. Signalling by the RuvB-like Pontin and Reptin
RT proteins.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF233278; AAF4341.1; -.
DR FLYBase; FBgn0040078; pontin.
DR InterPro; IPR003593; -.
DR SMART; SM00382; AAA; 1.
SQ SEQUENCE 456 AA; 50215 MW; D2FA09186A1E9B0E CRC64;

Query Match 72.6%; Score 1643.5; DB 5; Length 456;
Best Local Similarity 68.8%; Pred. No. 6.3e-87;
Matches 313; Conservative 74; Mismatches 67; Indels 1; Gaps 1;

QY 1 MRLEEVOSTSKORATHTHTKIGGLDANGMAIALAGFVGQAARAEAGIAMDITROK 60
DB 1 MKIEEVSTVETORLAASHVKGGLDEVGAAVSHAGVQKARERAGIAMDITROK 60
QY 61 MAGRAVLLAGPATGKTALALGIAOELSGKVPFCPMVGSSEYSEVKKTEVLMENFRRAI 120
DB 61 MAGRAVLLAGPATGKTALALGIAOELSGKVPFCPMVGSSEYSEVKKTEVLMENFRRAI 120
QY 121 GLRIKENKEVEYEGVELSPERASTGGVAKSISHTVLSIKYKGTQKLDSDSYDAL 180
DB 121 GLRIKENKEVEYEGVELSPERASTGGVAKSISHTVLSIKYKGTQKLDSDSYDAL 180
QY 121 GLRIKENKEVEYEGVELSPERASTGGVAKSISHTVLSIKYKGTQKLDSDSYDAL 180
DB 121 GLRIKENKEVEYEGVELSPERASTGGVAKSISHTVLSIKYKGTQKLDSDSYDAL 180
QY 181 IREKAVGDVITYEANSAGVAVKRGVCSFATEYDLAEAEYVPIPKGEVHKKEIYODVTL 240
DB 181 IREKAVGDVITYEANSAGVAVKRGVCSFATEYDLAEAEYVPIPKGEVHKKEIYODVTL 240
QY 241 HDLDAANAPOGGODILSLMGOMKPKRTETETKLRQEIKNVNRVIDEGIAELVGVLF 300
DB 241 HDLDAANAPOGGODILSLMGOMKPKRTETETKLRQEIKNVNRVIDEGIAELVGVLF 300
QY 301 IDEVHMLDIECTSYLNRALESPLSVIATNRGICNVRG-TDMSPHGIPVDLDRLYI 359
DB 301 IDEVHMLDIECTSYLNRALESPLSVIATNRGICNVRG-TDMSPHGIPVDLDRLYI 359
QY 360 IRRETYGPTMIOIATIRAQVEIEMDESILAYLGEIGQOTSLRHAIDLISPAVSVKTN 419
DB 360 IRRETYGPTMIOIATIRAQVEIEMDESILAYLGEIGQOTSLRHAIDLISPAVSVKTN 419
QY 361 IRTLLYSTADMGIITKRAQTEGLQLEENAFRLSEIGSSFLRAYAVOLLFPAHQMCKVN 420
DB 361 IRTLLYSTADMGIITKRAQTEGLQLEENAFRLSEIGSSFLRAYAVOLLFPAHQMCKVN 420
QY 420 GREKICKADLEEVSGIYLDKASSARLLOEQERYI 454
DB 420 GREKICKADLEEVSGIYLDKASSARLLOEQERYI 454
QY 421 GRMOISKDIEDVHSLFLDAKRSKHLSEKNKFM 455
DB 421 GRMOISKDIEDVHSLFLDAKRSKHLSEKNKFM 455

RESULT 8
ID 003940 PRELIMINARY; PRT; 463 AA.
AC 003940;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE HYPOTHETICAL 50.5 KDA PROTEIN.
RN RVB1.
OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCB1_Taxid=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Oliver K., Harris D.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Barrett B., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z48784; CAA88704.1; -.
DR SGD; S0002598; RVB1.
DR InterPro; IPR003593; -.
DR SMART; SM00382; AAA; 1.
KW Hypothetical protein.
SQ SEQUENCE 463 AA; 50453 MW; 93D6ECH7006AF0F2 CRC64;

Query Match 67.1%; Score 1517.5; DB 3; Length 463;
Best Local Similarity 65.7%; Pred. No. 1.1e-79;
Matches 286; Conservative 81; Mismatches 67; Indels 1; Gaps 1;

QY 14 RITHTHTKIGGLDANGMAIALAGFVGQAARAEAGIAMDITROKMGRAVLLAGPPA 73
DB 23 RITHTHTKIGGLDANGMAIALAGFVGQAARAEAGIAMDITROKMGRAVLLAGPPA 73
QY 74 TGTALALGIAOELSGKVPFCPMVGSSEYSEVKKTEVLMENFRRAIIGRIKENKEVEYEG 133
DB 83 TGTALALGIAOELSGKVPFCPMVGSSEYSEVKKTEVLMENFRRAIIGRIKENKEVEYEG 133
QY 134 EVELSPERASTGGVAKSISHTVLSIKYKGTQKLDSDSYDALIREKAVGDVITYE 193
DB 143 EVELSPERASTGGVAKSISHTVLSIKYKGTQKLDSDSYDALIREKAVGDVITYE 193
QY 194 EANSAGVAVKRGVCSFATEYDLAEAEYVPIPKGEVHKKEIYODVTLHDLDAANAPOGG 253
DB 203 EANSAGVAVKRGVCSFATEYDLAEAEYVPIPKGEVHKKEIYODVTLHDLDAANAPOGG 253
QY 254 QDILSLMGOMKPKRTETETKLRQEIKNVNRVIDEGIAELVGVLFIDEVHMLDIECT 313
DB 263 QDILSLMGOMKPKRTETETKLRQEIKNVNRVIDEGIAELVGVLFIDEVHMLDIECT 313
QY 314 YLNRALESPLSVIATNRGICNVRG-TDMSPHGIPVDLDRLYIIRRETYGPTMIO 372
DB 323 YLNRALESPLSVIATNRGICNVRG-TDMSPHGIPVDLDRLYIIRRETYGPTMIO 372
QY 373 ILAIRAQVEIEMDESILAYLGEIGQOTSLRHAIDLISPAVSVKTRGKICKADLEEV 432
DB 383 ILAIRAQVEIEMDESILAYLGEIGQOTSLRHAIDLISPAVSVKTRGKICKADLEEV 432
QY 433 SGILYLDKASSARLLOEQERYI 457
DB 443 KLFLLDARSTKILE 457

RESULT 9
ID 017607 PRELIMINARY; PRT; 458 AA.
AC 017607;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE C27H6.2 PROTEIN.
GN C27H6.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCB1_Taxid=6239;
RN (1)
RP SEQUENCE FROM N.A.

RA Gardner A.;
 RN Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: 3681042; CAB02793.1; -.
 DR InterPro: IPR003593; -.
 DR SMART: SM00382; AAA; 1.
 SQ SEQUENCE 458 AA; 50616 MW; F801E34FADCC0EE2 CRC64;

Query Match 57.9%; Score 1311; DB 5; Length 458;
 Best Local Similarity 57.8%; Pred. No. 8.3e-68;
 Matches 262; Conservative 78; Mismatches 91; Indels 22; Gaps 5;

OY 3 IEEVOSTSK-QRIATHHRIKIGLIDANGM-AIALAGFVGQAAREAGLAVDMIRQK 60
 DB 20 IEEVPTKQRIASHHVGIGIDTETQAHNEAGFVGQAARFRASTIVMIRKIC 79
 OY 61 MAGRAVLGAPPAKGTALALGIAOELGSKVPCPMGSEVYSEVKKTEVLMENFRRAI 120
 DB 80 MAGRAVLGAPPAKGTALALGIAOELGSDGVPFVPLVASEVSEVKKTEVLMENFRRAI 139
 OY 121 GLRIKENKEVEGETEELPEEAESTTGGYAKSISHVITSIKTKYKQKOLKDSITDAL 180
 DB 140 GLRIKENKEVEGETEELPEEAESTTGGYAKSISHVITSIKTKYKQKOLKDSITDAL 180
 OY 181 IEEVANGDVYIIEANSAGVAVKRCDFATEYDLEAEVYPIPKGVHKKKEVDVTL 240
 DB 199 LKQHEVEVDVYIIEANSAGVAVKRCDFATEYDLEAEVYPIPKGVHKKKEVDVTL 240
 OY 241 HDLDAANAPQGGG-DLISLMGQMKPRKTEITTEKLRQELINKVYNYRIDEGIAELVPGVL 299
 DB 259 HDLDAANAPQGGG-DLISLMGQMKPRKTEITTEKLRQELINKVYNYRIDEGIAELVPGVL 299
 OY 300 FIDEVHMLDIECSYLRALESPLSPVILATNRGICNVGTDWTSFPHGIPVLDRLVY 359
 DB 319 FIDEVHMLDIECSYLRALESPLSPVILATNRGICNVGTDWTSFPHGIPVLDRLVY 359
 OY 360 IRPFTYGTPTMIOQLARAQVEELIDMEESLAVLGIGQOTSLRHAQLISPAVSKTN 419
 DB 379 IPTKTYNEEDIRKILVHRTAEVNOFEKAFDLTRLCAGTC----- 420
 OY 420 GREKICKADLEEVSGYLIDAKSSARLLOEQER 452
 DB 421 GREVIEVEDVDRCTKLPMNGESLKAABEMRO 453

RESULT 10
 OY 094692 PRELIMINARY: PRT; 465 AA.
 AC 094692;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 GN SPB83.08.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;

OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=972;
 RA Wood V., Rajandream M.A., Barrell B.G., Bothe G., Pohl T.,
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO THE AAA FAMILY OF ATPASES.
 DR EMBL: AL035536; CAB36870.1; -.
 DR InterPro: IPR003593; -.
 DR SMART: SM00382; AAA; 1.
 DR KW Hypothetical protein; ATP-binding.
 FT NP_BIND 72
 SQ SEQUENCE 465 AA; 51561 MW; 150E10A1D31F847 CRC64;

Query Match 43.7%; Score 989.5; DB 3; Length 465;
 Best Local Similarity 46.9%; Pred. No. 2.6e-49;
 Matches 210; Conservative 87; Mismatches 136; Indels 15; Gaps 5;

OY 10 SKRQRIATHHRIKIGLIDANGMALALAGFVGQAAREAGLAVDMIRQKMAGRAVLLA 69
 DB 12 SKRQRIATHHRIKIGLIDANGMALALAGFVGQAAREAGLAVDMIRQKMAGRAVLLA 71
 OY 70 GPATGKTALALGIAOELGSKVPCPMGSEVYSEVKKTEVLMENFRRAIGRIKENKE 129
 DB 72 GPATGKTALALGIAOELGSDGVPFVPLVASEVSEVKKTEVLMENFRRAIGRIKENKE 131
 OY 130 YRGEVTELSPEEAESTTGGYAKSISHVITSIKTKYKQKOLKDSITDALJREKAVGD 189
 DB 132 YRGEVTELSPEEAESTTGGYAKSISHVITSIKTKYKQKOLKDSITDALJREKAVGD 189
 OY 190 VIYEANSAGVAVKRCDFATEYDLEAE--VPIPKGVHKKKEVDVTLHDLAAN 247
 DB 186 VIYEANSAGVAVKRCDFATEYDLEAE--VPIPKGVHKKKEVDVTLHDLAAN 245
 OY 248 ADPQGGODLISLMGQMKPRKTEITTEKLRQELINKVYNYRIDEGIAELVPGVL 307
 DB 246 ADPQGGODLISLMGQMKPRKTEITTEKLRQELINKVYNYRIDEGIAELVPGVL 307
 OY 308 DICEFSYLRALESPLSPVILATNRGICNVGTDWTSFPHGIPVLDRLVYIRTEYGP 367
 DB 299 DICEFSYLRALESPLSPVILATNRGICNVGTDWTSFPHGIPVLDRLVYIRTEYGP 367
 OY 368 FEMIOQLARAQVEELIDMEESLAVLGIGQOTSLRHAQLISPAVSKTN 427
 DB 359 FEMIOQLARAQVEELIDMEESLAVLGIGQOTSLRHAQLISPAVSKTN 427
 OY 428 DLEEVSGYLIDAKSSARLLOEQERYIT 455
 DB 419 DIRRYVELFLDQKRSEVYLEEYGRNVIT 446

RESULT 11
 OY 09DE27 PRELIMINARY: PRT; 462 AA.
 AC 09DE27;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 GN REPTIN.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
 OC NCBI_TaxID=8555;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Etard C., Wedlich D., Kuehl M.;
 RT "Xenopus laevis reptin, a Xenopus homolog of the prokaryotic RuvB
 RT helicase.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF218071; AAC44126.1; -.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Gohshi T., Shimada M., Kawahire S., Ichimura T., Omata S., Horigane T.;
 RT "Molecular cloning of mouse p47, a second group mammalian RuvB DNA
 RT helicase-like protein: homology with those from human and
 RT Saccharomyces cerevisiae.";
 RL J. Biochem. 0:0-0(1999).
 DR EMBL; AB013912; BAA/6297.1;
 DR MGD; MGI:1342299; Ruvb12.
 DR InterPro; IPR003593;
 DR SMART; SM00382; AAA; 1.
 KM Helicase.
 SQ SEQUENCE 463 AA; 5112 MW; 419D779972933308 CRC64;

Query Match 43.0%; Score 974; DB 11; Length 463;

Best Local Similarity 44.0%; Pred. No. 2e-48;

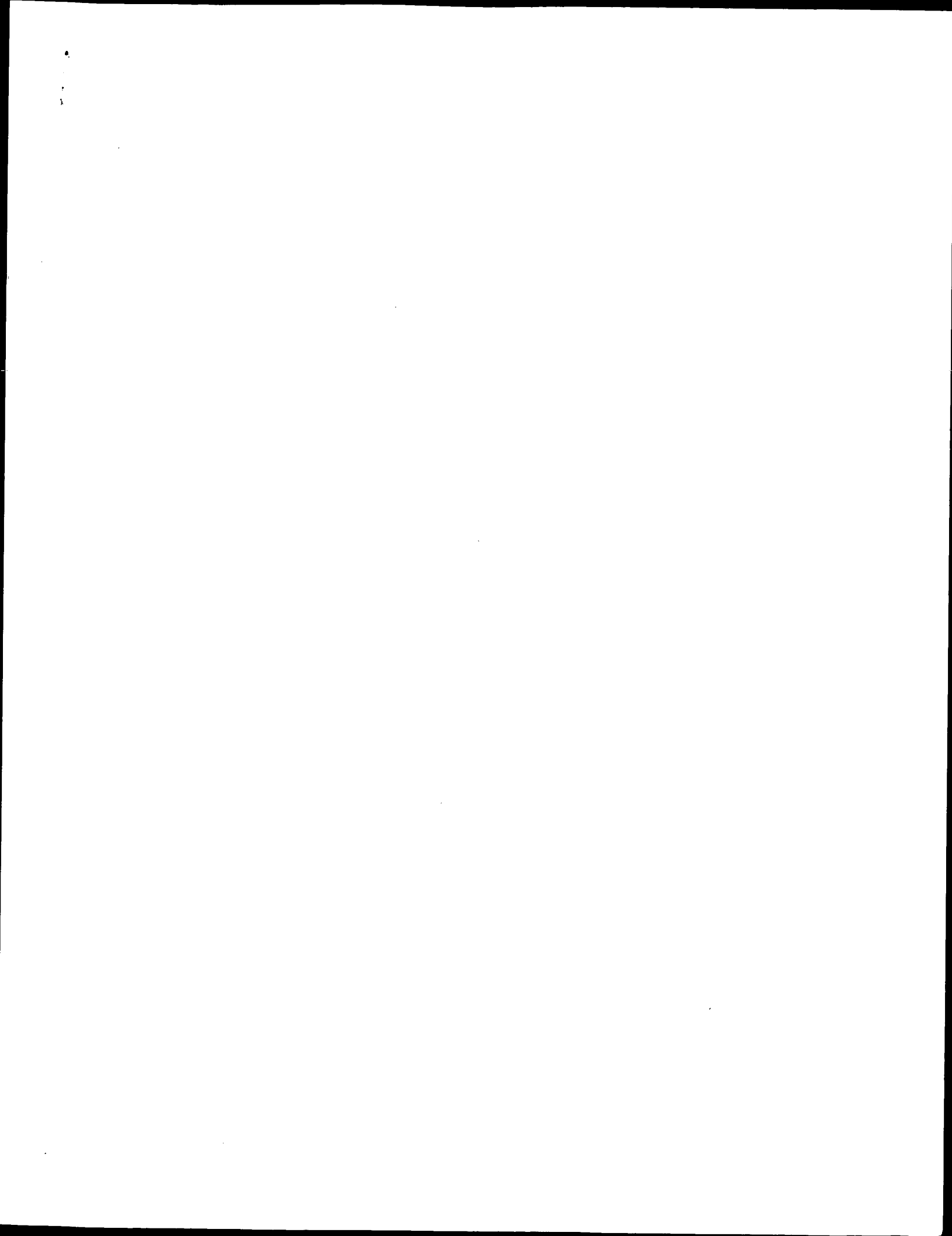
Matches 200; Conservative 97; Mismatches 144; Indels 14; Gaps 4;

```

OY 2 RIEVQSTSKQRIATHTTHIKGLIDANGMAIALAGFVGQAAREAGLAVDMIRQKM 61
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 9 KVEIRIDVTIERIGAHSHIRGLDLEPRQASQGVQIAARRAGVLEMIREGKI 68
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
OY 62 AGRVLLAGPATKKTALALGIAELGSKVPCPMVSGSEVSSSEVKTVELMENFRRAIG 121
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 69 AGRVLLAGPATKKTALALGIAELGSKVPCPMVSGSEVSSSEVKTVELMENFRRAIG 128
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 122 LRIKENKEVEEGETELSPPEAEESTTGGYANSISHVITSLKTVKTKQLKIDSSIYDALI 181
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 129 VRIKETETIEEGEVEIQTIDPATGTGSKVKG---LTLKTEMETIYDLGTMIESLT 183
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
OY 182 KERVAVGVDTIYIANSQAVKRYGRDPSFATEYDLAE--YVIPKGEVHKKEIVQDVT 239
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 184 KDVVOAGDVITTDKAKGKISKLGKSTFRARDYDAMGSGTKFYQCPDGELOKKEVYHTVS 243
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
OY 240 LHDLDANAQPOGQDILSLMGQMMKPRKTEITERKIROEINKVNVRYIDEGIAELVPGVL 299
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 244 LHEIDVINSRTQG--FLAL---FSGDTGEIKSEVREQINAKVAEMWREGKAEIIPGVL 296
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
OY 300 FIDEVHMLDIECESYINRALESPLSIYIATNRCICNVGTDMTSPHGIPVDLIDLRLVI 359
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 297 FIDEVHMLDIESFSLNRALESMDAPVLLIMATNRGITRIRGTSYOSPHGIPIDLRLVI 356
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 360 IRFETYGPTMIOIILAIRAOVEIDMEESLAYIGEIGOOTSLRHAIDLISPASYVSKTN 419
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 357 VSTSPYSEKDTKOILIRCEEEVEVMSDEDAYVLRIGLETSLRAYAQLITAAASLVCRRK 416
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
OY 420 GREKICKADLEEVYGLYLDKASSARLLQOQERYI 454
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 417 KGEVVOVDIKRYVSLFLDESRSSTQYMKKEYODAFI 451
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

Search completed: November 13, 2001, 07:02:32
 Job time: 353 sec




```

|||||
294 CCTGGCGGTCGCCGCCACGGCAAGCGCGCTAGCGCTGGCATAG 343
84 laingluuuglyserlyvalprophesyrometvalglysergiu 100
|||||
344 CCCAGGAGCTCGCGAGAGGTCCTTTCTGCTATGCTAGGATCAGAA 393
101 valtyrsersergiuvalylslythrvalleuemetgluasnphar 117
394 GTGACTACTCGGAGGTCAAGAAAGTGAAGTGTGATGAAAAATTTCCG 443
117 gtagalaleglyleuargllelysluasnlysgluvaltyrgluglyg 134
|||||
444 TAAGGCTAATAGGTTTGGCTAATAAGAAAAAAGAGGTTTATGAAGAG 493
134 luvalthrleuuserprogluvalagluuserthrthrlyglylythr 150
494 AGGTTACTGAACTTCCCGAAGAGGCTGAGAGTACAACTGGTGGATAT 543
151 alalyserlleserhisvallelleserleuylsvallysglythr 167
544 GCAAAAGATTAGCCATGATCATCATTAAGACTGTAAAGGAGC 593
167 tyrgluuileuylsleuaspserserlletyraspalaaleuileysglu 184
594 TAAGCAACTGAAGTTAGATTCTTCAATTATGATGCTGTATCAAGAAA 643
184 ysvalalavalyspvalyletyrileglualaaserglyvalval 200
644 AGGTGGCAGTGGGTGATGTTATATCATCGAAGCAATAGGAGCGAGT 693
201 lysarvalylslyrasyaspsersphenalathrlythrleuuglu 217
694 AAAAAGNTTGGTAGATGATTTCTTGTCAAGAAATACGATCTTGAAGC 743
217 agluuutyryvalproleprolysglygluvalhislyslslysglu 234
744 TGAAGAGATGTTCTTATCCCAAGGTGAAGTCCATTAAGAAAAAAGAA 793
234 levalgluaspvalthrleuuhisaspaleuaspalaalaasnaaglin 250
794 TTGTCAGAGATGCACATTCATGACCTGATGAGCAAAATGCTCAGCCA 843
251 glnglygluasnpiileuserleuemetgluemetleystproar 267
844 CAAGGTGGCCAAAGATTTTGTCCCTTATGCGCAGATGATGAACACG 893
267 glythrthgluilethrthgluylsleuargluuileasnlyvalval 284
894 AAAAGACTGAATCACCGAAAACTACGCCCAAGAAATTAAGTGTGTA 943
284 snatgytyrileaspgluyllealagluuvalproglyvalleuphe 300
944 ATGATATTCGATGAAGGATGACAGCTGTGTACTGTGTGTGTGTGTC 993
301 ileaspgluvalhismetleuasnpiileglucyspheserlyleuasn 317
994 ATTGATGAGTCCACATGTTGATGATGATGTTTCTTATCTTAACCG 1043
317 galaleuuglyserproleuserproilleuualathrasnarag 334
1044 TGCATTGGAGAGCCCATATACCAATCGTACTGTACTAATAATAGG 1093
334 lytlecysasnvalargglythrasspmetthserprohisglylepro 350
1094 GAATATGTAATGTAAAGGACTGATGACAAAGTCCACATGCTATACCG 1143
351 valaspaleuaspargleuvalilleleargthrlythrlythrlythr 367
1144 GTGGATCTTCTAGATAGGCTGTGATTAATGCGACAGACATATAGGCC 1193
367 orhrgrumetileglinleuualaleargalagluvalgluile 384
|||||

```

```

1194 TACTGAGATGATACAGATATTGGCTATCCGACACAAAGTGAAGATTTG 1243
384 spmetaspgluuglyserleuualalythrleuuglyluileglyluinthr 400
1244 ATATGATGTGAAGAAAGTCTTGTCTTATTTAGCGAATGCGACAGACACA 1293
401 Serleuarghisalaleglinleuileserproalaservaliserly 417
1294 TCTTTAGACATGCTATTCATGATGATTCACCTGCGACGCTGTCTCAAA 1343
417 sthrasnlyarggluylsillecyslysalaspseuugluvalserg 434
1344 GACTAATGAGAGAGAAATTCGACAGCTGATCTCAGAGAACTCAGTG 1393
434 lytleuylsleuaspalalysserseralargleuileugluinlin 450
1394 GGCTATATTGGATGCAAAATCTCGGCTGCTGCTCAGAGACAAACAA 1443
451 gluargtyrillethr 455
1444 GAAAGATACATCAC 1458

```

```

seq_name: gb_pat2:AX077224
seq_documentation_block:
LOCUS AX077224 1845 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 1 from Patent WO0105975.
ACCESSION AX077224
VERSION AX077224.1 GI:13121818
KEYWORDS
SOURCE
Zeae mays.
ORGANISM Zeae mays.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1845)
AUTHORS Mahajan,P.B.
TITLE Maize orthologues of bacterial ruvb:cdnas and uses thereof
JOURNAL Patent: WO 0105975-A 1 25-JAN-2001;
PIONEER HI-BRED INTERNATIONAL, INC. (US)
FEATURES
Location/Qualifiers
1..1845
/organism="Zea mays"
/db_xref="taxon:4577"
85..1452
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC32388.1"
/db_xref="GI:13121819"
/translation="MRLIEVOSTSKKQRTATHTTHIKGLGDANGMALAAGFVGOSA
AREAAAGLAVDMIRKKMAGRAVILNCPATGKTALADGIAOELGSKVPPCWGSEVY
SSEYKRTVELENERRAIGLRIRKNEKVEGEVETLSPEAESITGVAKSIISHYIS
LKYVKGKRLKLDSSITDVALIKERVAAGDVITYEANSQAVKRVGRDSEFATYDLEAE
EYVPIPKGEVHKKEIYVDVTLHDLDAAANQPOGODILSLMGOMMKPKRTETKLR
QEINKVNRVIDEIAELVPGVLFIDEVHLDIECFSYLRALLESPLIVTLATNRG
ICNVRGTDMSPHGIPVDLRLVITRTTEYGTETIOLATRAOVEDIDMEESLAY
LGEIGQOTSRHMIQLISPAVSYSKINGREKICKADLEESGLYDAKSSARLQEOQ
ERYTT"

```

CDs

```

BASE COUNT 530 a 378 c 490 g 445 t 2 others
ORIGIN

```

alignment_scores:

```

Quality: 2253.00 Length: 455
Ratio: 4.952 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.341

```

alignment_block:

```

US-09-589-510-4 x AX077224

```

```

Align seg 1/1 to: AX077224 from: 1 to: 1845

```

```

1 Metargillegluvalgluuserthserlyslslnargilealath 17

```

seq_name	gb_pat2	AX077228
985	ATTGACGAGTCCATCGATGTGGATATCGAATGTTTTTCTTATCTTAACCG	1034
317	gAlaLeuGluSerProLeuSerProIleValIleLeuAlaItrAsnArg	334
1035	TGCATTGGAGCCCATTTATCACCAAATTGGATVACTTGGTACGAATVAGG	1084
334	lYlleCysAsnValArgGlyThrAspMetHisSerProHisGlyLeuPro	350
1085	GAATATGTAATGTAAGAAGAACTGATATGACAGTCACATGSGTATACCA	1134
351	ValAspLeuLeuAspArgLeuValIleIleArgThrGluThrTyrGlyPr	367
1135	GGGATCTCTTGAATAGGTGGTGGTATTTTCGGACAGACACTATGGCCC	1184
367	onhrglmeTlleGlnIleLeuAlaIleArgAlaGlnValGluGluIleA	384
1185	TACTGATGATACAGATATTGGCTATCCGACGACAAAGTGGAGGACATTG	1234
384	spMetAspGluGluSerLeuAlaTyrLeuGlyGluIleGlyGlnGlnThr	400
1235	ATATGATGAAAGAAAGTCTGCTTATTTAGGCGGAGTCCGACGACGACA	1284
401	SetLeuArgHisAlaIleGlnLeuIleSerProAlaSerValValSerLy	417
1285	TCCTTTAGACATGCTATTCATTTGATATACCTCCAGCGTGGTCTCAAA	1334
417	sThrAsnGlyArgGluIuysIleCysLysAlaAspLeuGluGluValSerG	434
1335	GACTAATGCAAGAGAAATATGCAAGGCTGATCTCGAGGAAGTCAAGTg	1384
434	lyLeuTyrLeuAspAlaLysSerSerAlaArgLeuLeuGlnGlnGln	450
1385	GGCTCTATTTGGATGCCAATCTCGGCTCGGCTCCGAGGACAAACA	1434
451	GluArgTyrIleThr	455
1435	GAAGAATCATCTACC	1449
seq_name	gb_pat2	AX077228
seq_documentation_block	1886 bp	DNA
LOCUS	AX077228	1886 bp
DEFINITION	Sequence 5 from Patent WO0105975.	PAT
ACCESSION	AX077228	
VERSION	AX077228.1	GI:13121822
KEYWORDS		
SOURCE	Zea mays.	
ORGANISM	Zea mays	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
AUTHORS	Mahajan, P. B.	
TITLE	Maize orthologues of bacterial ruvB:cdnas and uses thereof	
JOURNAL	Patent: WO 0105975-A 5 25-JAN-2001;	
FEATURES	Plasmid: HI-BRD INTERNATIONAL, INC. (US)	
source	location/Qualifiers	
	1..1886	
	/organism="Zea mays"	
	/db_xref="taxon:4577"	
	82..1449	
	/note="unnamed protein product"	
	/codon_start=1	
	/protein_id="CAC32390.1"	
	/db_xref="GI:13121822"	
	/translation="WRIEVOSTSKKORITHTHIKIGLIDANGSMPLAAGFYGOAA	
	AREAGLADMTROKKKAGRALLAGPPATGKTALALGIAOELGSKVPCMGSEVY	
	SSVKKTEVLMENEFRAIGRIKIKENVEYGEVIELSPBAESTTGTAKSISHYIIG	
	LTVAGTOKLQDPSITYDALIKKEVAAVDVITYEANSAGVAKRGKDSFATIEDLAA	
	EVPYIPKGVKKKEIVODVTIHLDAANAOPOGGODITSLMGQMKPKETIEKLE	
	OEINIKVEVHYIDEQIAELVPGVFLIDEVHMLIDECFSYNRALESPLSYILATNR	
	ICAVRTDITSPHGCPVLDLDRIVITRTYGTETMIOLAIRAQVEEIDLEESIAA	
	LGRTGQTSIRHAIQILDPASVAKTNGREKMKADLEESVGLYDAKSSARILQEQ	

BASE COUNT 524 a 408 c 518 g 436 t
ORIGIN

alignment_scores:

Quality: 2222.00 Length: 455
Ratio: 4.916 Gaps: 0
Percent Similarity: 99.341 Percent Identity: 97.582

alignment_block:

us-09-589-510-4 x AX077228 ..

Align seg 1/1 to: AX077228 from: 1 to: 1886

```

1 MetArgllleGlulgluValGlnSerThrSerLysLysGlnArgllleAlaTh 17
  |||
82 ATGAGATGAGAGAGAGTGCATGACCTCGAAGAAAGCAGCGCATCGCCAC 131
  |||
17 rHlSTrHlSlleLysGlyLeuGlyLeuAspAlaasnGlyMetAlaIleA 34
  |||
132 CCACACTCACATCAAGGCGCTCGCGCTCGACGCCAATGGAATGTCGATGC 181
  |||
34 lalaValaIaaglyPheValIglValIaIaIaIaArgGluAlaIaagly 50
  |||
182 CGTTGGCGGGCGGGGTTCTGGGCGGCGGCGGCGGCGGCGGCGGCGG 231
  |||
51 LeuAlaValaIaSpMetIleArgGlnLysLysMetAlaIaArgAlaValle 67
  |||
232 CTGGCGGCTGCATGATCGCGCAGAAAGATGGCGGCTCGCGGCTGCT 281
  |||
67 uleuAlaIaIaGlyProAlaIaThrGlyLysThrAlaIaIaIaIaIa 84
  |||
282 CTTTGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 331
  |||
84 lalGlnGluLeuGlySerLysValProPheCysProMetValIaIaSerGlu 100
  |||
332 CCGAGAGAGCTCGCGCAGCAAGGCTCCATCTCTCTATGCTAGAGATCGAA 381
  |||
101 ValTYrSerSerGluValLysLysThrGluValLeuMetGluAsnPhar 117
  |||
382 GGTACTCTCCAGAGGTCAGAAAGAGAGGCTGATGGAATATTCGCG 431
  |||
117 GARAlaIleGlyLeuArgIleLysGluAsnLysGluValTYrGluGly 134
  |||
432 TAGACCTTAGCTTGGCTATTAAGAAACCAAGAGGTTATGAAGGAG 481
  |||
134 luValThrGluLeuSerProGluGluIaIaIaIaIaIaIaIaIaIa 150
  |||
482 AGTTATGAGACTTCCCGAAGAGAGGCTGAGAGCAACAAGTGGTGAAT 531
  |||
151 AlalysSerIleSerHisValIleIleSerLeuLysThrValLysGlyTh 167
  |||
532 GCGAAAGACATTAGCCACGATATCATTTGGCTTAAGACTGTCAAGGAG 581
  |||
167 rLysGlnLeuLysLeuAspSerSerIleTYrAspAlaIaIaIaIaIa 184
  |||
582 TAACCAATTGAANTAGACCCCTCATTTATGCTTGTGATCAAGGAA 631
  |||
184 ysaValaIaValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 200
  |||
632 AGGGGAGAGGAGGAGTGTATATACATGAAGCAAAATAGTGAGCAGAG 681
  |||
201 LysArgValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 217
  |||
682 AAAAGAGTGGTAGATGCTATCTTTTGGTACAGATATGATCTTGAAG 731
  |||
217 agLugluTYrValProIleProLysGlyGluValHisLysLysGlyLut 234
  |||
732 TGAAGAGATATGTTCTATCCCAAGGTAAGTCCATTAAGAAAGGAA 781
  |||
234 lvalaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 250
  |||

```

```

782 TAGTCAGAGATGTCACACTCCATGACCTTGATGACAGCAAAATGCCAGCA 831
  |||
251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetMetLysProAr 267
  |||
832 CAAGGTGGCCAGATATTTTGTCTTATGGCCAGATGATGATGAGCCAG 881
  |||
267 glyThrGluIleThrGluLysLeuArgGlnLutIleAsnLysValaIa 284
  |||
882 GAAGACTGAATATCCCGAAAGCTACGCCAAGAAATCAATAGTGTGATA 931
  |||
284 snArgTYrIleAspGluGlyIleAlaGluLeuValProGlyValLeuPhe 300
  |||
932 ACAGATATATCGACGAGAGATCGAGAGCTTGTACCTGGGTGTTTGTTC 981
  |||
301 IleAspGluValHisMetLeuAspIleGluCysPheSerTYrLeuAsnAr 317
  |||
982 ATTGATGAGGTCCACATGTTGATATTGAATGCTTTTCTTATCTTAACG 1031
  |||
317 glAlaLeuGluSerProLeuSerProIleValIleLeuAlaThrAsnArg 334
  |||
1032 TGCATTGGAGAGCCCATTAATCCAAATTGTGATACCTCCCTACGAATGAG 1081
  |||
334 lYLecYsaValaIaArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 350
  |||
1082 GAATATGTATATGTGAGAGAACCGATATGACGAGTCCACATGTATACCA 1131
  |||
351 ValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 367
  |||
1132 GTGGACCTTCTAGATAGCTTGGTGTATTTGCGACAGAAACATATGCCCC 1181
  |||
367 oThrGluMetIleGlnIleLeuAlaIleArgAlaGluValGluGluIleA 384
  |||
1182 TACTGACATGATACGATACGATCTGCTATCCGAGCACAAGTGAAGAGATTG 1231
  |||
384 sPmetAspGluGluSerLeuAlaTYrLeuGlyLutIleGlyGlnIaThr 400
  |||
1232 ATATCGATGAAGAAAGCTTCTTATTAGCGAGATCGGACGACGACACA 1281
  |||
401 SerLeuArgHisAlaIleGlnLeuIleSerProAlaSerValIaIaIa 417
  |||
1282 TCTTTGACACATGCTATTTCAGTTCTCATACCTGCGCCAGCGCTGCGCAAA 1331
  |||
417 sThrAsnGlyArgGluLysIleCysLysAlaAspLeuGluGluValaIaSer 434
  |||
1332 GACCAAGCGGAGAGAAAGATGTCCAAAGGCTGACCTCGAGAGAGATCAGG 1381
  |||
434 lYleuTYrLeuAspAlaLysSerSeraIaArgLeuLeuGlnGlnGln 450
  |||
1382 GGCCTATTTTGGATGGCCAAATCTCGGCTGCGCTGCTCCAGAGGACACA 1431
  |||
451 GluArgTYrIleThr 455
  |||
1432 GAAAGATACATCAC 1446
  |||
seq_name: gb_pat2:AX077230
seq_documentation_block:
LOCUS AX077230 1898 bp DNA PAT 22-FEB-2001
DEFINITION Sequence 7 from Patent WO0105975.
ACCESSION AX077230
VERSION AX077230.1 GI:13121824
KEYWORDS
SOURCE
ORGANISM
Zea mays.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 1898)
AUTHORS
Mahajan,P.B.
Maize orthologues of bacterial rnpB:cdnas and uses thereof
Patent: WO 0105975-A 7 25-JAN-2001.
JOURNAL
PIONEER HI-BRED INTERNATIONAL, INC. (US)
Location/Qualifiers

```


source 1. 1898
/organism="Zea mays"
/db_xref="taxon:4577"
166. .1536
CDS /note="unnamed protein product"
/protein_id="CAC32391.1"
/db_xref="GI:13121825"
/translation="MARIEVOSTSKORIRATHTHIKGLIDQANGSMPLAAGFVGA
AAREAGLAVDMIRKRMAGRALLLAGPRATKALAGIQLGSGKVPQPMQSEV
YSEEVKTEVLEMFERRAIGLRKENKEVEEVEEVEEVEEVEEVEEVEEVEE
GLKTVGTOLKLDPSIDALIKEKVAADVITYEENSVAVRGRCDSFRTEYLEA
EEVYIPKCEVHKKEIVODVTLHDIDANAPQOGDILSLMGQMRPKTEITKL
ROENKVNREYIDEGIAELVPGVLEIDEVHMDIECFYLNLSPLSPYIATLNK
GICNVGTQMTSPHGIPLDRLVIRIETGPTFMIOILIRAVEEIDIDESLA
YLGIRGQTSRLRAIQLLSASVAVATNREKMKADLEVSGLVLAKSSARLIQEQ
QERTY"

BASE COUNT 532 a 413 c 521 g 432 t
ORIGIN

alignment_scores:
quality: 2211.50 Length: 456
Ratio: 4.893 Gaps: 1
Percent Similarity: 99.123 Percent Identity: 97.368

alignment block:
US-09-589-510-4 x AX077230 ..

Align seq 1/1 to: AX077230 from: 1 to: 1898

```

1 MetArgIleGluGluValGlnSerThrSerLysLysGlnArgIleAlaThr 17
166 ATGAGGATAGAGGAGGTGCAATCGACCTCGAAGAACACGCGATCGCCAC 215
17 RHSTHRIHISILelysGlyLeuGlyLeuAsp...AlaasnGlyMetAlaI 33
216 CCACACCTACATCAAGGCGCTCGCGCTCGACGCGCCATGCAATGTGCA 265
33 leaIalaGluAlaIaGlyPheValGlyAlaAlaAlaIaargGluAlaIa 49
266 TGCCGTTGGCGGCGGCTTCTGTCGCGAGCGCGCGCGCGCGCGCGCC 315
50 GlyLeuAlaValaIaSpMetIleArgGlnLysLysMetAlaGlyArgAla 66
316 GGGGTGGGCGTGCACATGATCCGCGAAGAACATGCGCGTGGCGCGCT 365
66 lleuLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyI 83
366 GCTCCTTCGGGGCCGCCGCCGCGCGCAAAACGCCCTGACCGCTCGGCA 415
83 leaIaGlnGluLeuGlySerLysValaProPheCysProMetValaGlySer 99
416 TAGCCGACGAGCTCGGCGAGCAAGGTCCCATTCGTCTATGTAGATCA 465
100 GluValaIYrSerSerGluValaLysLysThrGluValaLeuMetGluAsnPh 116
466 GAAGGTACTCTCTCAGAGGTCAAGAAACGTGAGTGTGATGAAAAATTT 515
116 eArgArgAlaIleGlyLeuArgIleLysGluAsnLysGluValaIYrGln 133
516 CCGTAGAGCTATAGGTTTGGCTATTAAGAAACAAAGAGTTATGAAG 565
133 IYgIuValaIThrGluLeuSerProGluGluAlaGluSerThrThrGlyGly 149
566 GAGAGGTATTTGAACTTCCAGAGAGGCTTGAGAGCACAACTGGGTGA 615
150 TyrAlaLysSerIleSerHisValaIleIleSerLeuLysThrValaLysG 166
616 TATGGCAAAAGCATTAAGCACCTATTCATTCGTTAAAGACTGTCAAGG 665
166 yThrLysGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysG 183

```

```

666 GACTAAGCAATTGAATTAAGACCTTCATTTATGATGCTTGTGATCAAG 715
183 IuLysValaIaIaValaGlyAspValaIleIYrIleGluAlaAsnSerGlyAla 199
716 AAAAGTGGCAGTGGTGATCTTATATACATGAAGCAAAATAGGGAGCA 765
200 ValLysArgValaGlyArgCysAspSerPheAlaThrGluTyrAspLeuG 216
766 GTGAAACAGAGTTGGTAGTGTGATTCCTTTCTGCTACAAATATGATCTGA 815
216 uAlaGluGluTyrValaProIleProLysGlyGluValaHisLysLysG 233
816 ACCTGAGAGGTATGTTCTTATCCCAAGGTGAAGTCCATTAAGAAAAAG 865
233 IuIleValaGlnAspValaIThrLeuHisAspLeuAspAlaAlaAsnAlaGln 249
866 AATATGTGCGAGATGTCACTCATGACCTTGATGACGCAAAATGCCAG 915
250 ProGlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLysPr 266
916 CCACAAAGTGGCCAGATATTTTGTCCCTTATGCGCCAGATGATGAAGCC 965
266 oArgLysThrGluIleThrGluLysLeuArgGlnGluIleAsnLysVala 283
966 ACGAAGACTGAATCACCCAAAGCTACGCGCAAGAAATCAATAGGTGG 1015
283 alaSnArgTyrIleAspGluGlyIleAlaGluLeuValaProGlyValaLeu 299
1016 TAAACATATATATGACGAGAAAGATGCGAGACTGTACCTGCTGTGTG 1065
300 PheIleAspGluValaHisMetLeuAspIleGlyCysPheSerTyrLeuAs 316
1066 TTCATTGATGAGGTCCACATGTTGATATTTGAATGCTTTTCTTATCTTAA 1115
316 naArgAlaLeuGluSerProLeuSerProIleValaIleLeuAlaThrAsn 333
1116 CCGTGCACTTGAGAGGCCATTATACCAATGTGATACCTGGTACAGATA 1165
333 rGlyIleCysAsnValaIargGlyThrAspMetThrSerProHisGlyIle 349
1166 GAGCAATATATATGTGAGAGAACCGATATGACGATCGACATGGTATA 1215
350 ProValaIspLeuLeuAspArgLeuValaIleIleArgThrGluThrTyrS 366
1216 CCACTGACCTTCTAGATAGGTTGGTCAATATTCGACAGAAACATATGG 1265
366 yProThrGluMetIleGlnIleLeuAlaIleArgAlaGlnValaGluGlu 383
1266 CCTTACTGAGATGATACAGATACCTGCTATCCGAGCACAAGTGAAGAGA 1315
383 leaSpMetAspGluGluSerLeuAlaIYrLeuGlyGluIleGlyGlnGln 399
1316 TTGATATCATGAGAAAGACCTTCTTATTTAGCGGAGTGGACAGCAG 1365
400 ThrSerLeuArgHisAlaIleGlnLeuIleSerProAlaSerValaIse 416
1366 ACATCTTTGAGACATGCTATTACAGTGTGCTATACCTCCAGCGGTGGCG 1415
416 rIYsThrAsnGlyArgGluLysIleCysLysAlaAspLeuGlnGluVala 433
1416 AAAGACCAAGCGAGAGAAAGATGTGCAAGGCTGACCTGAGAGAACTCA 1465
433 eArgLysThrLeuAspAlaLysSerSerAlaArgLeuLeuGlnGlnGln 449
1466 CGGGCTCTATTTGGATGCCAAATCTCGCTCGCTGCTCCAGAGACAA 1515
450 GlnGluArgTyrIleThr 455
1516 CAAGAAAGATATCATCACC 1533

```

seq_name: gb_pat2:AX077232

seq_documentation_block:

LOCUS	AX077232	1869 bp	DNA	PAT	22-FEB-2001
DEFINITION	Sequence 9 from Patent WO0105975.				
ACCESSION	AX077232				
VERSION	AX077232.1 GI:13121826				
KEYWORDS	Zea mays.				
SOURCE	Zea mays.				
ORGANISM	Zea mays.				
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.				
AUTHORS	Mahajan, P. B.				
TITLE	Male orthologues of bacterial ruvB:cdns and uses thereof				
JOURNAL	Patent: WO 0105975-A 9 25-JAN-2001;				
FEATURES	PIONEER HI-BRED INTERNATIONAL, INC. (US)				
source	location/Qualifiers				
	1..1869				
	/organism="Zea mays"				
	/db_xref="taxon:4577"				
	64..1380				
	/note="unnamed protein product"				
	/codon_start=1				
	/protein_id="CAC32392.1"				
	/db_xref="GI:13121827"				
	/translation="MRIEVOYSTSKRORIATHTHIKIGLDANGMALATAAAGFYGKK MAGNAVLAGPATGKTLAAGIAOELSKVPCFPMWSEVSEYSPFKTEVIMENERRA IGLILKENKEYEDEVTELSPEEAESTTGAKSISHTVLSIKTKTKOLKLDSDTY DALIKENKEYEDEVTELSPEEAESTTGAKSISHTVLSIKTKTKOLKLDSDTY ODVPLHLDANAQPOGCDILSLMGOMAKPKREITEKLRQETNKVNYR IDEGAE LVPEVLPIDEYHMLDIECFSTLNALDESPISPLVLTATNGICNVRTDMTSPHGIPV DLDRIVYIRREITGPEMTOLIAIRAQVEYIDMDESLAYLGEIGQOTSLRAIQLI SPASVSKTNGREITCAADLEEVSGVLYLDAKKSARLLQEOERYIT"				
BASE COUNT	539 a 373 c 486 g 471 t				
ORIGIN					
alignment_scores:					
Quality:	2157 50				
Ratio:	4.926				
Percent Similarity:	96.264				
Length:	455				
Gaps:	2				
Percent Identity:	96.264				
alignment_block:					
US-09-589-510-4 x AX077232	..				
Align seg 1/1 to: AX077232 from: 1 to: 1869					
1 Metatglliegugluvalglnserthserlyssglnarlglaealath 17					
64 ATGAGGATCGAGGAGGTGCACCTCGAAGACGACGCCATCGCCAC 113					
17 rhisrthistilelysgyleuglyleuasplalsnglymetalailea 34					
114 CCACACCACATCAAGGACTCGGCTCGACGCCAATGAGATGCGCATG 163					
34 lalauaialaaglyphevalglynalalaalaararglualaalagly 50					
164 CGTGGCGGCGGCGGTCTGTGGC..... 186					
51 leualvalaspmetilearnglnlysllymetalaiglyaralavalle 67					
187CAGAAAGAAATGCGCGCGCGCGGTGCT 215					
67 ulauaialglyproproalathrarglystraralauaialaueuglyilea 84					
216 CCTTGGCGGCTCGCGCCGCCAAGGCAAGCGCCCTAAGC...GGCATG 262					
84 laingluileuglyserlyvalprophcysprometvalaiglysergu 100					
263 CCACGAGGCTCGGCGACCAAGGTCCCTTTCGTGCTCATGTAGATCAGAA 312					
101 valtyrsersergluvalylslstrngluvalleumetcluasnphear 117					
313 GTGTACTCTCGAGAGTCAAGAAACTGAGGTGCTGATGGAATAATTTCG 362					

[illegible]

417 sThAsnGlyArgGluuylleCysLysAlaAspLeuGluGluValSerG 434
 |||
 1263 GACTAAATGAGAGAGAAAATCTGCAAGCTGATCTCGAGAGAACTAGTG 1312
 434 LLeuTYrLeuAspAlaLysSerSerAlaArgLeuLeuGlnGln 450
 |||
 1313 GGCTCTATTGGATGCCAATCCCTCGGCTGCTCCAGAGCAACAA 1362
 451 GLeuTYrLeuThr 455
 |||
 1363 GAAGATACATCACC 1377
 seq_name: gb_D13:CAR276264
 seq_documentation_block:
 LOCUS CAR276264 1755 bp mRNA PLN 06-MAR-2000
 DEFINITION Cicer arletinum mRNA for putative Ruv DNA-helicase.
 ACCESSION AJ276264
 VERSION AJ276264.1 GI:7208770
 KEYWORDS putative Ruv DNA-helicase.
 SOURCE chickpea
 ORGANISM Cicer arletinum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Cicer.
 1 (bases 1 to 1755)
 Dopic, B., Esteban, R. and Labrador, E.
 A RuvB DNA-helicase like protein is expressed in chickpea epicotyls
 Unpublished
 2 (bases 1 to 1755)
 Labrador, E.
 Direct Submission
 Submitted (29-FEB-2000) Labrador, E., Dpto. Fisiologia Vegetal,
 Univ. Salamanca, Campus Miguel de Unamuno. Pza. Doctores de la
 Reina s/n, E-37007, SPAIN
 Location/Organism
 1. 1755
 /organism="Cicer arletinum"
 /cultivar="Castellana"
 /db_xref="taxon:3827"
 /country="Spain"
 /dev_stage="seedling"
 /tissue_type="etiolated epicotyls"
 /clone_lib="CAN-5"
 /clone="CanRuv"
 /note="age 5 days"
 207. 1583
 /codon_start=1
 /product="putative Ruv DNA-helicase"
 /protein_id="CAB76908.1"
 /db_xref="GI:7208771"
 /translation="MEKKIEVOSTTKKORAVTHTHIKIGLGEVSKALPASPFGV
 CAEAREACGLVVDIMROKMKAGKALLLAGPGTKTALAGICQELSTKVPFCPMGVS
 EYVSEVKEKTEVLEMEFRRAIGRIKENKEVEVEVLESPETESLTGGYGSISHV
 IIGKTVPGTKOLKPTIDYDALIKEKAVADVITIEANGAVKRVRSAPATEPDL
 EAEKVPKPKGVKKKEIYODVTLHDLDANAPOGODILSLMGMMKPRKTEITD
 KLRQETNKVKNRYIDEGYAEVLPVCLTIDEVHMLDMCFESYLNALSSSPYIVFAT
 NNGICTVRSQTDTSFGILPVDLDRLVYIIRQTGPAEIIQIIILAROVEIIVDEES
 LAFIEIGQRTSLRHAQVLSPASVVAIKINGDNICAKADLEICSLILDKSSAKILLO
 EOEKRYIS"
 BASE COUNT 525 a 332 c 410 g 488 t
 ORIGIN
 alignment_scores:
 Quality: 2039.00 Length: 455
 Ratio: 4.582 Gaps: 0
 Percent Similarity: 97.802 Percent Identity: 87.473
 alignment_block:
 US-09-589-510-4 x CAR276264 ..
 Align seg 1/1 to: CAR276264 from: 1 to: 1755

1 MeTArgIleGluGluValGlnSerThrSerLysLysGlnArgIleAlaTh 17
 |||
 216 ATGAAATAGAGAGAGTTCAATCCACGACGAGAAACACAGGACTTCTAC 265
 17 rHisThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleA 34
 |||
 266 TCACACTACACTAAAGCCTAGAGTCTCGAGAGTGCAGTGAAGCATTCG 315
 34 LLeuAlaAlaGlyPheValGlyGlnAlaAlaAlaArgGluAlaAlaGly 50
 |||
 316 CTTTTCATCTGATTTGTTGGGCAAGCTGAGGCACAAAGAGCATGTGGG 365
 51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaValle 67
 |||
 366 CTTGTGTGATATGATATGCTGACAGAGAGATGGCTGCAAGCAGCTCC 415
 67 LLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
 |||
 416 TCTCCGCGGCGCACCTGGGACCGCAAGACTGCTTAGCAGCTGGGATAT 465
 84 LLeuGlnLeuGlySerLysValProPheCysPrometValGlySerGlu 100
 |||
 466 GTCAGAGCTTGGACCAAGGTTCCATTTGCCCAATGTGATAGTTTCGAA 515
 101 ValTYrSerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
 |||
 516 GTATATTTCACAGAAAGTAAAGACGAGGTCTTATGAAATTTTCG 565
 117 GArgAlaIleGlyLeuArgIleLysGlnAsnLysGluValTYrGlnGlyG 134
 |||
 566 ACGGCTATTGGTCTACGATTAAGGAAATTAAGGAAATATATAGAGGCG 615
 134 LValTYrThrLysLeuSerProGluGluAlaGluSerThrTYrGlyTYr 150
 |||
 616 AGGTGACTAGCTCTCTCCGAAAGACTGAGAGTTTAACAGAGGTAT 665
 151 AlaLysSerIleSerHisValIleIleSerLeuLysThrValLysGlyTh 167
 |||
 666 GGTAAAGATTATAGCCAGCTGAATATGGTGAAGACAGTTAAAGAGAC 715
 167 rLysGlnLeuLysLeuAspSerSerIleTYrAspAlaLeuIleLysGlu 184
 |||
 716 CAAGCAACTAAAGTTGACCCCAAGATATATATGATGCTTGAATAGGAAA 765
 184 YsValAlaValGlyAspValIleTYrIleGluAlaAsnSerGlyAlaVal 200
 |||
 766 AGGTAGCTGTGGGAGATTTATATATATAGGCAAAATAGCGGGCTGTG 815
 201 LysArgValGlyArgCysAspSerPheAlaThrGluTYrAspLeuGluVal 217
 |||
 816 AAAAGGGTTGGCCGAAGTGAATGCTTTTGTACGATTTGACCTTGAACG 865
 217 aGluGluTYrValProIleProLysGlyGluValHisLysLysGlyIle 234
 |||
 866 AGAGAGATGATGTTCCACTTCTTAAGGAGAGGTTTCACAAAAAGAGAGA 915
 234 LLeuAlaGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnPro 250
 |||
 916 TTGTTCAGAGATGTAAACCTACATGATCTGATGCTGCAAGCAGACACT 965
 251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetMetLysProAr 267
 |||
 966 CAAGGGGGCAAGATATTTTGTCTTATATGAGCCAGATGATGAAGCTAG 1015
 267 GLeuThrGluIleThrGluLysLeuArgGlnGluIleAsnLysValAla 284
 |||
 1016 GAAAGACAGAAATCACTGACAGATTGAGACAAAGAAATTAATAGGTTGTC 1065
 284 sNArgTYrIleAspGluGlyIleAlaGluLeuValProGlyAlaLeuPhe 300
 |||
 1066 ACCGATATATGATGAAGGTGTTCAGAGCTTGTCTCGGGGTTCTATTT 1115

CDS	/gene="TIP49"	54..1424	/function="a putative DNA helicase"
	/gene="TIP49"		/codon_start=1
	/product="TIP49"		/protein_id="BAA28169.1"
	/db_xref="gi:3132308"		/translation="MKIEEVKSTKTQRIASHSHVKGIGLIL
			AREAGAGVLEILSKMAGRAVILLAGSPGGKTLALAIAC
			STELKTEVLEMFENFRALIGLRKTEKVEEVEYELPCE
			LKTKAGKGLKIDPSIESPSLOKEKEVAGDYIYEANGAVK
			EYVFLPGDVHKKEIITQDVTLDHLDVNAARPOGGDILSM
			GEIKRVNKKYITDQSIAGELVPGVFLVEVDHMIIDIEFYLR
			NCVIRGTEIDTSPGIPLDLRVMIITRTMLPYPOEKQITIR
			HLEIGIKTILRYSVQLTFPANILAKINGKDSIEKEHVEEI
			ODKKMK"
			1710
polya_site	/note="11 a nucleotides"		
BASE COUNT	458 a	396 c	472 g 384 t
ORIGIN			
alignment_scores:	Quality: 1723.50	Length: 455	
	Ratio: 4.113	Gaps: 1	
Percent Similarity:	92.088	Percent Identity:	72.527
alignment_block:			
US-09-589-510-4 x	AB012122	..	
Align seg 1/1	to: AB012122	from: 1	to: 1710
1	MetatgllleglugluvalaGlnSerThrSerlyslslnargillea	atrh 17	
	: : : : : : : :		
54	ATGAGAGATTGAGGAGGAGGAGACACACGACGACGACGACGATCGCTC	109	
17	RhstThrslllelsgslyleuuglyleuaSpalaaSnglymetala	llea 34	
	: : : : : : : :		
104	CCACAGCCACGATGAAGAGCGCTGGGGCGTGGAGCAGAGCGCGCTTGCC	AAGC 153	
34	laleualaalaaglyphevalaGlylnalaalaalaarguaalaagly	50	
	: : : : : : : :		
154	AGCGGGCGCTCAGGGCTTGTGGCCAGAGAACCGCGAGAGGCATGTGGC	203	
51	leualaValaSpmetlleargGlnlyslslymetalaGlyargala	valle 67	
	: : : : : : : :		
204	GTCATAGTACGAATTATTCAAACACCAAGAAAATGCTGGGAAGACTGTCTT	253	
67	uleualaGlyProProalaThrGlyThrAlaLeualaLeuGlyIleA	84	
	: : : : : : : :		
254	GTTGGAGAGACCTCCCGAAGACTGGCAAGACACTGCTGCGCATTTG	303	
84	laglgluenuuglyserlyslsValProphicysPrometValaGly	Serglu 100	
	: : : : : : : :		
304	CTCAGAGAGCTGGGTAGTAAAGTCCCTTCGCCAATGCTGGGAGTGAA	353	
101	ValTyrSerSerglValAllyslsThrGluValLeuMetGluAsp	Phear 117	
	: : : : : : : :		
354	GTTTACTCACTGAGATCATCAAGAACAGAGGCTGATGAGGAACCTTCG	403	
117	gargalailegyleuAarglllelysluAsnlysluValTyrGlu	glugly 134	
	: : : : : : : :		
404	CAGGGCCATTGGGCGCAATTAAGAGAACCAAGAAATTATGAAGTG	453	
134	luvalThgluLeuSerProgluGluAlaGlnSerThrThhGlyGly	Tyr 150	
	: : : : : : : :		
454	AAGTCACAGACACTTACCTCGTGTAGACAGAGCAATCCCATGGAGGATAT	503	
151	AlaYsSerIleSerHisValIleIleSerleuYsthrValYsls	GlyTh 167	
	: : : : : : : :		
504	GCGCAAAACCAATTAGCCATGTGATCATGAGGATCAAAACCACTGCAAGGAC	553	

```

167 rlysglnleuLysLeuAspSerSerIleTyrAlaLeuIleLysGluL 184
|||||
554 CAACAGCTTGAACACTGACCCAGCATTTTGAAGTTTGCAAGAAAGAC 603
|||||
184 ysvAlaValAlaGlyaspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
|||||
604 GAGTAGAAGCTGAGATGTGATTTACATTTGAAGCCAAACAGTGGGGCGCTG 653
|||||
201 LysArgValGlyArgCysAspSerPheAlaThrGlyTyrAspLeuGluL 217
|||||
654 AAGAGCAGAGGCGAGGTGTATACCTTATGCCACAGAAATTCGACCTTGAAAC 703
|||||
217 aGluGlyTyrValProIleProLysGlyLysValHisLysLysGlyL 234
|||||
704 TGAAGGTATGTCCCTTGCCAAAGGGATGTGCACAAAAGAAAGAA 753
|||||
234 lValAlaGlnaspValThrLeuHisAspLeuAspAlaAlaAsnIleGlnPro 250
|||||
754 TCATCCAAAGATGTGACCTTGACATGACTTGCTTAATGCGCGGCC 803
|||||
251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLysProAr 267
|||||
804 CAGGGGGGCAACAGATATCTGTCCATGATGGCGCAGCTATGAAAGCCAA 853
|||||
267 GlySerThrGluIleThrGlyLysLeuArgGlnGluIleAsnLysValAla 284
|||||
854 GAAGACAGAAATCAGACAACTCGAGGGGAGATTAAATAGTGGTGA 903
|||||
284 snArgTyrIleAspGlnGlyIleAlaGlnLeuValProGlyValLeuPhe 300
|||||
904 ACAAGTACATCGACAGGCGATGCTGAGCTGCTCCGGGTGCTGTT 953
|||||
301 IleAspGlnValHisMetLeuAspIleGluCysPheSerTyrLeuAsnAr 317
|||||
954 GTTGATGAGGCGCACATGTGACATGATGAGTCTGCTCACTGACCG 1003
|||||
317 GAlaLeuGlnSerProLeuSerProIleValIleLeuAlaThrAsnArg 334
|||||
1004 CGCCCTGAGCTCTGATCGCTCCATCGTATCTTTCATCCAAACCGAG 1053
|||||
334 LylCysAsnValArgIleYThr...AspMetThrSerProHisGlyIle 349
|||||
1054 GCAACTGTGTATCAGAGGACCTAGAGACATCAATCCCTCAGCGCATC 1103
|||||
350 ProValAspLeuLeuAspArgLeuValIleIleArgThrGluThrYrGI 366
|||||
1104 CCTCTTGACCTTGTGACCGAGTATGATATCCGAGCATGCTGTATAC 1153
|||||
366 yProThrGlnMetIleGlnIleLeuAlaIleArgAlaGlnValGluGluI 383
|||||
1154 TCCACAGGAATGAACAGATCATTAATAATCCGTCAGAGGAGAA 1203
|||||
383 LeaspMetAspGlnGlnSerLeuAlaTyrLeuGlnGlyIleGlyGln 399
|||||
1204 TCACATCAGAGGAGGAGGACACTGAACACCTGGGGAGATTGCGACCAAG 1253
|||||
400 ThrSerLeuArgHisAlaIleGlnLeuIleSerProAlaSerValIse 416
|||||
1254 ACCACACTGAGTACTCAGTGCACCTGCTGCCCGGCAACTGTGCTTC 1303
|||||
416 rLysThrAsnGlyArgGlnLysIleCysLysAlaAspLeuGlnValS 433
|||||
1304 TAAATTCACAGGGAAGACACATTTGAGAAAGACATCTGCAGACATCA 1353
|||||
433 eTgIleuTyrLeuAspAlaLysSerSerAlaArgLeuLeuGlnGln 449
|||||
1354 GTGACTTTTCTATGATGCCAAGTCTCCGCCAAATCTCGTGACAG 1403
|||||
450 GlnGlnArgTyrIle 454
|||||
1404 CAGGATTAAGTACATG 1418
|||||
seq_name: gb_pat2:E21816

```

```

seq_documentation_block: 1730 bp DNA PAT 07-FEB-2001
LOCUS E21816
DEFINITION Protein forming a complex with TBP, polynucleotide encoding said
protein, antisense polynucleotide against said polynucleotide, and
antibody recognizing said protein.
ACCESSION E21816.1 GI:13023712
VERSION JP 1999009285-A/2.
KEYWORDS unclassified.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1730)
AUTHORS Toshihiko, K.T.T. and Makino.
TITLE Protein forming a complex with TBP, polynucleotide encoding said
protein, antisense polynucleotide against said polynucleotide, and
antibody recognizing said protein
JOURNAL Patent: JP 1999009285-A 2 19-JAN-1999;
SUMITOMO ELECTRIC IND LTD
COMMENT OS Unidentified
PN JP 1999009285-A/2
PD 19-JAN-1999
PF 27-JUN-1997 JP 1997187398
PR
PI TOSHIHIKO, KISHIMOTO, TAKAKI TAMURA, YASUTAKA MAKINO PC
C12N15/09,A61K48/00,C07K14/47,C07K16/18,C12P21/02,C12P21/08, PC
C12Q1/68
PC C12N15/00
CC Topology: Linear; Location/Qualifiers
FH Key 1..1730
FT source /organism='unclassified'.
FEATURES
source Location/Qualifiers
1..1730
/organism='unclassified'
/db_xref='taxon:32644'
BASE COUNT 469 a 401 c 476 g 384 t
ORIGIN
alignment_scores:
Quality: 1723.50 Length: 455
Ratio: 4.113 Gaps: 1
Percent Similarity: 92.088 Percent Identity: 72.527
alignment_block:
US-09-589-510-4 x E21816 ..
Align seg 1/1 to: E21816 from: 1 to: 1730
1 MetArgIleGluGlnValGlnSerThrSerLysLysGlnArgIleAlaThr 17
|||||
60 ATTAAGATTTGAGGAGGTGAAGACACTAGAGACGACGCGATGCGCTC 109
|||||
17 rHisThrHisIleLysGlyLeuGlnLysLeuAspAlaAsnGlyMetAlaIleA 34
|||||
110 CCACAGCCACAGTGAAGGGCTGGGCTGAGACGAGACGCGCTGGCCAAAC 159
|||||
34 lAlaLeuAlaIleGlyPheValGlyGlnAlaAlaAlaIleArgGlnAlaIleGly 50
|||||
160 AGCGGGCTCTAGGGCTTGTGGGCCAGAGACAGCGGAGAGCGATGTGGC 209
|||||
51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyValArgAlaValIe 67
|||||
210 GTCATAGTAGAATTAATCAAAACAGAAATGTGTAAGACCTGCTCTT 259
|||||
67 uLeuAlaGlyProProAlaThrGlnLysThrAlaLeuAlaLeuGlyIleA 84
|||||
260 GTTGGCAGGACCTCTGGAACTGCAAGACAGCTGAGCTTGCTTGCTATTG 309
|||||
84 lArgGlnGluLeuGlySerLysValProPheCysProMetValGlySerGlu 100
|||||
310 CTCAGAGACTGGGTAGTAAGTCCCTTCTGCCCAATGTGTGGGAGTGA 359
|||||

```

	400	ThrserrLeArghIsAlalIeglIneuIsesrProAlaservAlvalSe	416
	1260	ACCACACTGAGTACTCACTGACGTGGTGACCCCGGCACAATTCCTGCC	1309
	416	rlysrhrnsglYArgGlulysIlleCylsylaAlaspIeugluValS	433
	1310	TAAATTCACAGGGGAAGCACCATTTGACAAAAGCATGTCGAAGAATCA	1359
	433	erglYleuTYrLeuAspAlaLysSerSerAlaArgLeuencIngIn	449
	1360	GTAACCTTTTCTATGATGTCACCAAGTCTCTCCGCCAATAATCTGCTGAC	1409
	450	glnGluatgTYrlle	454
	1410	CAGGATAGTACATG	1424
seq_name:	gb_pr8: HSA010058		
seq_documentation_block:			
LOCUS	HSA010058	1733 bp	mRNA
DEFINITION	Homo sapiens mRNA for NMP238 protein.	PRI	20-NOV-1998
ACCESSION	AJ010058		
VERSION	AJ010058.1 GI:3892583		
KEYWORDS	NMP238 protein.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
JOURNAL	Holzmann,K. Submitted (03-AUG-1998) Holzmann K., University of Vienna, Institute of Tumor Biology-Cancer Research, Borschkegasse 8a, Vienna, A-1090, AUSTRIA 2 (bases 1 to 1733) Holzmann,K., Gerner,C., Korosec,T., Polt,A., Grimm,R. and Sauermann,G. Identification and characterization of the ubiquitously occurring nuclear matrix protein NMP 238 Biochem. Biophys. Res. Commun. 252 (1), 39-45 (1998)		
FEATURES			
source	Location/Qualifiers		
	1..1733	/organism="Homo sapiens"	
	/db_xref="taxon:9606"		
	/clone="IMAGE_591284"		
	/tissue_type="pancreas"		
	77..1447	/function="Related to AAA family, role in ATP-dependent anchoring"	
	/codon_start=1		
	/product="Nuclear matrix protein NMP238"		
	/protein_id="CAA08986.1"		
	/db_xref="GI:3892584"		
	/translation="MKIEEVSTTKTORIASHSVKGLIDESGLAOASGLVGQENAREACGVIVELIKSKMAGRAVLALAPGTEKTLALAIADELCSKYFCFPGSEVYVSTELEKTETVMENFRRAIGLRKEKEVEEVELPCETENMGYGKTSIVIIIGLTKAKGOLKLIDPISIELOKEREVEDVIYIANGSGAVKRGRCRPVATEPLAEAYLVLPKGVDKKKEIIDDTVLHPDVAANRPOGGCILSMGOLMCKPKETDKLRGEINLVKNKYIDOGIAELVPGLFPVDEVHMLDICPTLYHRALLESSLAPIYIFRSNRGNVCVRGREDITSPHGIFLDLLDRMTIRIMLTPEOMKIITKAOTEGNISSEALNDHQDKTKTURLYSVOLTLPANLAKINGDSIERKEHEVEISELFYDAKSASKILLDOOQDKRMK"		
	1733	/note="18 a nucleotides"	
BASE COUNT	462 a	403 c	481 g 387 t
ORIGIN			
poly_a_site			
alignment_scores:			
Quality:	1723.50	Length:	455
Ratio:	4.113	Gaps:	1

alignment_block:
US-09-589-510-4 x HSA010058 ..

Align seg 1/1 to: HSA010058 from: 1 to: 1733

```

1 MetArgIleGluGluValGlnSerThrSerIleuArgIleAlaThr 17
   ::::::::::::::::::::::::::::::::::::::::::::::
77 ATGAAGATTGAGAGGTAGACAGCACTACGAAGACGCGCATCGCTC 126
   ::::::::::::::::::::::::::::::::::::::::::::::
17 HisThrIleIleuGlyLeuGlyLeuAspAlaAsnGlyMetAlaIle 34
   ::::::::::::::::::::::::::::::::::::::::::::::
127 CCACAGCCAGCGTGAAGGGCTGGGGCTGGAGCAGCGCTTGCGCAAGC 176
   ::::::::::::::::::::::::::::::::::::::::::::::
34 LeuAlaAlaGlyPheValGlyGlnAlaAlaIleArgIleAlaIle 50
   ::::::::::::::::::::::::::::::::::::::::::::::
177 AGCGGGCTCAGGGCTGTGGGCCGAGAGAAACGCGCAGAGCGATGTGC 226
   ::::::::::::::::::::::::::::::::::::::::::::::
51 LeuAlaValAspMetIleArgIleIleuIleuSerMetAlaGlyAlaVal 67
   ::::::::::::::::::::::::::::::::::::::::::::::
227 CTCATAGTAGAATTATATCAAAAGCAAGAAATGCTGGAAGAGCTGTCT 276
   ::::::::::::::::::::::::::::::::::::::::::::::
67 uLeuAlaGlyProProlaThrGlyLysThrAlaLeuAlaLeuGlyIle 84
   ::::::::::::::::::::::::::::::::::::::::::::::
277 GTTGGCAGAGACCTCTGGAACCTGCACAGACAGCTCTGCTGCTATTTG 326
   ::::::::::::::::::::::::::::::::::::::::::::::
84 IAGIleGluLeuGlySerLysValProPheCysPrometValGlySer 100
   ::::::::::::::::::::::::::::::::::::::::::::::
327 CTCAGAGACCTGGTAGTAGGTCCCTTCTGCCAATGTTGGGGAGTGA 376
   ::::::::::::::::::::::::::::::::::::::::::::::
101 ValTyrSerSerGluValIleLysLysThrGluValIleuMetGluAsn 117
   ::::::::::::::::::::::::::::::::::::::::::::::
377 GTTTCCTCACTAGATGATCAAGACAGACAGAGTGCATGAGCAACTTCG 426
   ::::::::::::::::::::::::::::::::::::::::::::::
117 GARGAlaIleGlyLeuArgIleLysGluAsnLysGluValTyrGlu 134
   ::::::::::::::::::::::::::::::::::::::::::::::
427 CAGGGCCATTGGGCTGCAGATTAAGAGACCAAGAACTTATGAGAGTG 476
   ::::::::::::::::::::::::::::::::::::::::::::::
134 JuValThrGluLeuSerProGluGluAlaGlySerThrThrGlyGly 150
   ::::::::::::::::::::::::::::::::::::::::::::::
477 AAGTACAGAGCTAACTCCGTGTGAGACAGAAATCCCATGGGAGATAT 526
   ::::::::::::::::::::::::::::::::::::::::::::::
151 AlaLysSerIleSerHisValIleIleSerLeuLysThrValLysGly 167
   ::::::::::::::::::::::::::::::::::::::::::::::
527 GCGAAACCATTAAGCATGTGATCTAGACTCAAAACAGCCAAAGAGAC 576
   ::::::::::::::::::::::::::::::::::::::::::::::
167 rLysGluLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLys 184
   ::::::::::::::::::::::::::::::::::::::::::::::
577 CAACAGATTGAACCTGACCCGACCATTTTGAAGTTTGCAGAAAGAGC 626
   ::::::::::::::::::::::::::::::::::::::::::::::
184 yValAlaValAlaGlyAspValIleTyrIleGluAlaAsnSerGlyAla 200
   ::::::::::::::::::::::::::::::::::::::::::::::
627 GAGTGAAGACTGGAGATGTGATTACATTGAGCCACAGAGTGGGGCTG 676
   ::::::::::::::::::::::::::::::::::::::::::::::
201 LysAlaValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGlu 217
   ::::::::::::::::::::::::::::::::::::::::::::::
677 AAGAGCAGAGGCGAGGTGTGATPACTATGCGCACAGATTCGACCTTGA 726
   ::::::::::::::::::::::::::::::::::::::::::::::
217 aGluGluTyrValProIleProLysGlyGluValHisLysLysGly 234
   ::::::::::::::::::::::::::::::::::::::::::::::
727 TGAAGAGTATGTCCCTTGCACAAAGGAGATGTGCACAAAGAAAGAA 776
   ::::::::::::::::::::::::::::::::::::::::::::::
234 LeuAlaGlyAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGln 250
   ::::::::::::::::::::::::::::::::::::::::::::::
777 TCATCCAAAGATGTGACTTGTGCTGACTTGAGTGGCTAAATGGCGGCG 826
   ::::::::::::::::::::::::::::::::::::::::::::::
251 GInGlyGlyGlnAspIleLeuSerLeuMetGlyIleMetLysProAr 267
   ::::::::::::::::::::::::::::::::::::::::::::::
827 CAGGGGGGAGCAAGATATCTGTCCATGATGGCCGCAATATGAAGCCAA 876
   ::::::::::::::::::::::::::::::::::::::::::::::
267 gLysThrGluIleThrGluLysLeuArgGlnGluIleAsnLysValAla 284
   ::::::::::::::::::::::::::::::::::::::::::::::
877 GAAGACAGAAATCACAGACAAACTCGAGGGAGATTTAATAGGTGTGA 926

```

```

284 snArgTyrIleAspGluGlyIleAlaGluLeuValProGlyValLeuPhe 300
   ::::::::::::::::::::::::::::::::::::::::::::::
927 ACAAGTACATCGACACGAGGCAATTGCTAGCTGCTCCCGGGTGTCTTT 976
   ::::::::::::::::::::::::::::::::::::::::::::::
301 IleAspGluValHisMetLeuAspIleGluCysPheSerTyrLeuAsn 317
   ::::::::::::::::::::::::::::::::::::::::::::::
977 GTTGATGAGGTCCACATGCTGAGCATTTGAGTGTCTTACCTACTTCAGCC 1026
   ::::::::::::::::::::::::::::::::::::::::::::::
317 gAlaLeuGlySerProLeuSerProIleValIleLeuAlaThrAsnArg 334
   ::::::::::::::::::::::::::::::::::::::::::::::
1027 CGCCCTGGAGCTTCTTATCGCTCCCATCGTCATCTTGTGATCCACCGAG 1076
   ::::::::::::::::::::::::::::::::::::::::::::::
334 LylIleCysAsnValArgGlyThr...AspMetThrSerProHisGly 349
   ::::::::::::::::::::::::::::::::::::::::::::::
1077 GCAACTGTGTATCATCAGAGGACCTGAGGACATCATCCCTCAGCGCATC 1126
   ::::::::::::::::::::::::::::::::::::::::::::::
350 ProValAspLeuLeuAspArgLeuValIleIleArgThrGluThrArg 366
   ::::::::::::::::::::::::::::::::::::::::::::::
1127 CCTCTTGACCTTCTGAGCCGAGTGTATGATATCCGACCACTGCTATAC 1176
   ::::::::::::::::::::::::::::::::::::::::::::::
366 yProThrGluMetIleGlnIleLeuAlaIleArgAlaGlnValGluGlu 383
   ::::::::::::::::::::::::::::::::::::::::::::::
1177 TCCACAGAAATGAAACAGATCATTAATAATCCGTGCCACAGCGAAGGAA 1226
   ::::::::::::::::::::::::::::::::::::::::::::::
383 LeuAspMetAspGluGlySerLeuAlaTyrLeuGlyGluIleGlyGln 399
   ::::::::::::::::::::::::::::::::::::::::::::::
1227 TCACATCATGATGAGAGGACCTGACACCTCGGGGAGATGGCACCAG 1276
   ::::::::::::::::::::::::::::::::::::::::::::::
400 ThrSerLeuArgHisAlaIleGlnIleLeuIleSerProAlaSerVal 416
   ::::::::::::::::::::::::::::::::::::::::::::::
1277 ACCCACTGAGGTATCATGATGATGATGATGATGATGATGATGATGAT 1326
   ::::::::::::::::::::::::::::::::::::::::::::::
416 rLysThrAsnGlyArgGluLysIleCysLysAlaAspLeuGluValAs 433
   ::::::::::::::::::::::::::::::::::::::::::::::
1327 TAAATACACGGGAGGAGACAGATGAGAAAGAGCATGTGAGAGATCA 1376
   ::::::::::::::::::::::::::::::::::::::::::::::
433 erGlyLeuTyrLeuAspAlaLysSerSerAlaArgLeuGluGlnGlu 449
   ::::::::::::::::::::::::::::::::::::::::::::::
1377 GTGAACTTTTCTATGATGCCAAGTCTCTCCGCAAAATCCTGGCTGAC 1426
   ::::::::::::::::::::::::::::::::::::::::::::::
450 GInGluArgTyrIle 454
   ::::::::::::::::::::::::::::::::::::::::::::::
1427 CAGGATTAAGTATCATG 1441
   ::::::::::::::::::::::::::::::::::::::::::::::

seq_name: gb_ptr4:AF099084

seq_documentation_block:
LOCUS      AF099084          1736 bp      mRNA          13-JAN-1999
DEFINITION Homo sapiens Pontin52 mRNA, complete cds.
ACCESSION  AF099084
VERSION    AF099084.1  GI:4151524
KEYWORDS   human.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1736)
            Bauer,A., Huber,O. and Kemler,R.
            Pontin52, an interaction partner of beta-catenin, binds to the TATA
            box binding protein
            Proc. Natl. Acad. Sci. U.S.A. 95 (25), 14787-14792 (1998)
JOURNAL    MEDLINE
            99061945
REFERENCE  2 (bases 1 to 1736)
            Bauer,A., Huber,O. and Kemler,R.
            Direct Submision
            Submitted (19-OCT-1998) Department of Molecular Embryology,
            Max-Planck Institute of Immunobiology, Stuebweg 51, Freiburg
            D-79108, Germany
FEATURES   location/Qualifiers
            source          1..1736
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"

```

```

/cell_line="SM480"
/tissue_type="colorectal adenocarcinoma"
65.1495
/function="binds to the TATA-box-binding protein (TBP)"
/codon_start=1
/product="Pontins2"
/protein_id="AAD04427.1"
/db_xref="GI:4151525"
/translation="MKIEVASTTKTORISHSHVKGGLGDESLAKOASGLVGQEN
AREQSVIVELIKSKMAGRAVLVLAGPGETKIALALAIQELSGKVPFCMVGSSEVY
STELKTEVLEMNENRRAIGLRKEITEYBEEVELPCEENPMGYGKTI SHVILIG
LTKAKGKQLKLADPSIESLSQKEVEAGDYI IANGSVAVRGRCPTVATEPFLAEH
EYVLPKGVDRKKEI IODVTLHDLDVANAPQGGDILSMGLMKPKTEIDKLR
GEIKVNVKYIDOGIAELVEPFLVDVHMDICDFYIHLAESSTAPVYIFASNGG
NCVIRGTEDTSPHCITPLDLDRMTITRMLYTPOEMKOITIKTAQTEGINISEALN
HGEIGTKTTLRTSVQLTPANLAKINGKDSIEKHEVEITSELYDAKSSAKTLADQ
QDKYMK"
BASE COUNT      473 a      400 c      481 g      382 t
ORIGIN
alignment_scores:
    Quality: 1723.50      Length: 455
    Ratio: 4.113      Gaps: 1
Percent Similarity: 92.088      Percent Identity: 72.527
alignment_block:
US-09-589-510-4 x AF099084 ..
Align seg 1/1 to: AF099084 from: 1 to: 1736
1 MetArgIleGluGluIuValGlnSerThrSerIlySlysGlnArgIleAlaThr 17
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
65 ATGAAGATTGAGGAGGTGAAGAGCACTACGAGAGACGACGCGATCGCTCC 114
17 rHsMrHsIleIlySgIlyLeuGlyLeuAspAlaSnGlyMetAlaIleA 34
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
115 CCACAGCCACGTAAGAGGGCTGGCGTGGACGACGAGCGCTGGCCACGC 164
34 IalEuAlaIaGlyPheValGlyGlnAlaAlaAlaArgGluAlaIaGly 50
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
165 AGCGCGCTCAGGGCTGTGGGCCAGAGAACGCGCGAGAGCAATGTGGC 214
51 LeuAlaValAspMetIleArgGlnIlySlysMetAlaGlyValAlaValle 67
::: |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
215 GTCATAGTGAATTAATCAAAAGCAAGAAAGCGCTGGAGAGAGCTGCTT 264
67 uEuAlaGlyProProlaThrGlyLysThrAlaLeuAlaGluIleA 84
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
265 GTTGGCAGGACCTCTCGGAGCTGGCAAGACAGCACTCTGGCTTGGCTATT 314
84 IalGlnIleuGlySerIlyValProPheCysProMetValIlySerGlu 100
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
315 CTCAGGAGCTGGGTAGTAGTCCCTTCTGCCCAATGCTGGAGGTGA 364
101 ValIlyrSerSerGluValIlySlysThrGluValIleuMetGluAspPheAr 117
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
365 GTTTCCTCAACTGAGATCAAGAAAGACAGAGGCTGATGGAACATTTCCG 414
117 gArGAlaIleGlyLeuArgIleIlySgluAsnIlySgluValTyrGluGly 134
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
415 CAGGGCCATTGGGCTGGCAATTAAGAGACCAAGAGAAAGTTTATGAAGGG 464
134 IuValIthGluLeuSerProGluGluAlaGluSerThrThrGlyGlyTyr 150
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
465 AAGTCACAGAGCTAACTCCGTGTGAGAGCAAGAAATCCCATGGAGAGATAT 514
151 AlalysSerIleSerHisValIleIleSerLeuIlySThrValIlySgluTh 167
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
515 GGCAAAACCATTAAGCATGTGATCATAGCACTCAAAACAGCCAAAGGAGAC 564
167 IlySgluIleuIlySleuAspSerSerIleTyrAspAlaIleuIlySglu 184
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

[illegible]

LOCUS	BC002993	1746 bp	mRNA	PRI	16-MAR-2001
DEFINITION	Homo sapiens, RuvB (E coli homolog)-like 1, clone MGC:3762, mRNA,				
ACCESSION	BC002993				
VERSION	BC002993.1	GI:12804268			
KEYWORDS	MGC.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 1746)				
JOURNAL	Strausberg,R. Direct Submission Submitted (06-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: DCTD/DTP cdNA Library Preparation: Rubin Laboratory cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada Info@cgsc.bc.ca				
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjelli, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lilja Prabh, Parvaneh Saeedi, Jacqueline Schehl, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyerdunyn, Marco Marra.				
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Series: IRAL Plate: 9 Row: k Column: 14. Location/Qualifiers 1..1746 /organism="Homo sapiens" /db_xref="LocustID:8607" /db_xref="taxon:9606" /clone="MGC:3762 IMAGE:2823568" /tissue_type="Lung, small cell carcinoma" /clone_lib="NH_MGC_7" /lab_host="DH10B-R" /note="Vector: pOTB7" 72..1442 /codon_start=1 /product="RuvB (E coli homolog)-like 1" /protein_id="AAH02993.1" /db_xref="GI:12804268" /translation="MKIEVYSTRKTQRIASHSHVKGLIGDESLAKQAASGLVGQENT AREAGCIVELIKSKMARAVILLAGPGGKTLALAIADQLSGKVPFCPMGSEEVY STELTKTEVMENERRAIGLRKETKEVEGEVELPCCETENMGGYGTI SHVIIGEGE LTKATGKOLKLADIPSEIQEKREVEAGVDYLANSAGVAKRGRCPTVATEPELEAE EYVLPKGDYHKKKETIIDDTLHLDPANRAPGGGDIILSMGDLMPKRTETDKLAR GEINKVANKYITDGIAELVPEVLVDFVDEHMILDTCFYLHRALESSTAPLVTFASNGG NCVIRGTEDITSPPGILDLDRVMI LRIMLYTPOEMKLIK I RAQREGINI SEALN HLGIGIKTRTLRYSVQLLTSPANLLAKINGKDSIEKHVEISELFDKSSAKILLADOO ODKYMK"				
CDS					
BASE COUNT	479 a	403 c	479 g	385 t	
ORIGIN					
alignment_scores:	Quality: 1723.50	Length: 455			
	Ratio: 4.113	Gaps: 1			
Percent Similarity:	92.088	Percent Identity:	72.527		
alignment_block:					

[illegible]

```

|||||
922 ACAAGTATCATGACACGAGCATGCTGAGCTGCCGGGTGTGCTTT 971
301 ILeaspIuValHisMetLeuAspIleGluCysPheSerTyrluAsnAr 317
972 GTTGATAGGTCCACATGCTGCATTTAGATGCTTACCTACCTGACCG 1021
317 gAlaLeuGluSerProLeuSerProIleValIleLeuAlaThrAsnArg 334
1022 CGCCCTGGAGCTCTTATCGCTCCCATCGCATCTTGTGCATCCACCGAG 1071
334 lylIecysAsnValArgGlyThr...AspMetThrSerProHISglyIle 349
1072 GCACACTGTGATCAGAGGACAGCAGACATCATCCCTCAGGCGATC 1121
350 ProValAspLeuAspArgLeuValIleIleArgThrGluThrTyrc 366
1122 CCTCTTGACCTCTGGACGAGATGATATCCGACCATGCTGTATAC 1171
366 yProThrGluMetIleGlnIleLeuAlaIleArgAlaGlnValGluGlu 383
1172 TCCACAGGGAATGAAACAGATCATTAATCCCTGCCAGACGAGAGAA 1221
383 lEaspMetAspGluSerLeuAlaTyrluGlyGluIleGlyGlnGln 399
1222 TCACATCATGAGGAGGACACTAACCCCTGGGAGATGGCACCAAG 1271
400 ThrSerLeuArgHisAlaIleGlnIleSerProAlaSerAlaValse 416
1272 ACCACACTGAGTACTCATGCTGCTGCAGCCGCCACACTGCTTGC 1321
416 lYsThrAsnGlyArgGluIleCysLysAlaAspLeuGluValS 433
1322 TAAATCAACGGGAGACAGCATGTGAAAGCATGTGCAAGAGATCA 1371
433 eArgIleuTyrluAspAlaLysSerSerAlaArgLeuLeuGlnGln 449
1372 GTAACTTTTCTATGATGCCAGTCTCCGCCAATAATCTGCTGACAG 1421
450 GlnGluArgTyrlle 454
1422 CAGGATAGTACATG 1436

seq_name: gb_pr4:AF070735
LOCUS AF070735 1750 bp mRNA PRI 17-NOV-1998
DEFINITION Homo sapiens RuvB-like protein RUVBL1 mRNA, complete cds.
ACCESSION AF070735
VERSION AF070735.1 GI:3243034
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 1750)
AUTHORS Qiu,X.-B., Lin,Y.-L., Thome,K.C., Pian,P., Schlegel,B.P.,
Weremowicz,S., Parvin,J.D. and Dutta,A.
An eukaryotic RuvB-like protein (RUVBL1) essential for growth
J. Biol. Chem. 273 (43), 27786-27793 (1998)
98447618
REFERENCE
2 (bases 1 to 1750)
Qiu,X.-B., Lin,Y.-L., Thome,K.C., Pian,P., Schlegel,B.P.,
Weremowicz,S., Parvin,J.D. and Dutta,A.
Direct Submission
Submitted (05-JUN-1998) Pathology, Brigham & Women's Hospital,
Harvard Medical School, 75 Francis Street, Thorn 630, Boston, MA
02115, USA
FEATURES
source 1..1750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"

```

```

CDS
/mmap="3q21"
77..1447
/codon_start=1
/product="RuvB-like protein RUVBL1"
/protein_id="AAC7819.1"
/db_xref="GI:3243035"
/translaton="MKIEVKSTTKTORIASHSVKGILDESGLAQKQAASGLVGOEN
ARACGVIVELIKSKMAGRAVLVLAGPGGRTALALAIQELGSKVPEPCPMVGSERY
STRIKREYVEMENFRRAIGRIKREKEVEGEVTELTPECTENPMGKYKTSYIIG
LTKAKTKQKLDPSIFESIQKRVAVGYITIANSCAVKRGKRDYATFELEAE
EYPLRFGDVHKKEIIOVYTLHDLDVANAARQGDILSMGQLMKPKRTITDLE
GELINKVNRKIDGIAELVGVLEDEVMHLDIECTYLDRALESIAIVIFASNRG
NCVIRGTEIDTSPHGIPDLDRVMIIRFMYTPQEMKQIIRIKRAQTEGINSEBAL
HGEISTKTITLRSVOLLTPANILAKINKOSIEKHYVEISELFPDAKSSAKIILADQ
ODKYMK"
BASE COUNT 476 a 403 c 484 g 387 t
ORIGIN
alignment_scores:
Quality: 1723.50 Length: 455
Ratio: 4.113 Gaps: 1
Percent Similarity: 92.088 Percent Identity: 72.527
alignment_block:
US-09-589-510-4 x AF070735
Align seg 1/1 to: AF070735 from: 1 to: 1750
1 MetArgIleGluGluValGlnSerThrSerLysLysGlnArgIleAlaThr 17
77 ATGAAGATTGAGAGGAGTGGAAGAGCATACGAAGACGACGCGCATCGCTTC 126
17 rHisThrHisIleLysGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleA 34
127 CCACAGCCACGTGAAGGGCTGGGGCTGGACGAGAGGCTTGGCCAAAGC 176
34 lAlaAlaIleAlaGlyPheValGlyGlnIleAlaIleArgGlnAlaIle 50
177 AGGGCGCTCTGAGGCTCTGTGGCCAGAGAAAGCGGACAGGCGATGTGGC 226
51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaValle 67
227 GTCATAGTAGAATTAATCAAAAGCAAGAAATGGCTGGAAGACCTGCTT 276
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaIleAlaIleGlyIleA 84
277 GTTGGCAGAGACTCTCTGGAAGCTGGCAAGACAGCTGTGCTGTGGCTATTG 326
84 lAglGlnLeuGlySerLysValProPheCysProMetValGlySerGlu 100
327 CTCAGGAGCTGGGTAGTAAGGTCCCTTCTGCCCAATAGTGGGAGATGAA 376
101 ValTyrlSerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
377 GTTACTCAACTGAGATCAAGAAGACAGAGAGTGTCTGAGAACTTCCG 426
117 gArgAlaIleGlyLeuArgIleLysGlnLysGlnLysGlnValTyrlGlnGly 134
427 CAGGCGCATTTGGGCTGGCAATTAAGGAGACCAAGGAAGTTTAAGAGGTG 476
134 lUValThrGluLeuSerProGluAlaGluSerThrThrGlyGlyTyrl 150
477 AAGTCACAGAGCTAAGTCCGTGTGAGACAGACAGATATCCATGGAGATAT 526
151 AlalysSerIleSerHisValIleIleSerLeuLysThrValLysGlyTh 167
527 GCCAAACCATTAAGCCATGTGATCATAGAGCTCAAAAGCCAAAGCAAGAC 576
167 rLysGlnLeuLysLeuAspSerSerIleTyrlAspAlaLeuIleLysGlu 184
577 CAACAGTTGAACCTGAGACCCGACATTTTGAAGTGTTCAGAAAGAGC 626

```



```

|||||
297 GTTGGCAGAGCTCTGGAAGTGGCAGACAGCTTGCTGCTGCTATG 346
84 lAgInGluLeuGlySerLysValProPheCysPrometValGlySerGlu 100
347 CTCAGAGAGCTGGGTAGTAGTCCCTTCTGCTCCCATGTGGGGAGTGA 396
101 ValTyrSerSerGluValLysLysThrGluValLeuMetGluAsnPhar 117
397 GTTACTACTAGATCAGACAGACAGAGGCTGATGAGAACTTCCG 446
117 gArGalaIleGlyLeuArgIleLeuGluAsnLysGluValTyrGlyG 134
447 CAGGGCCATTGGCTGCGAATAAGAGACCAAGAGATTATGAGGTG 496
134 lValThrGluLeuSerProGluGluValGluSerThrGlyGlyTyr 150
497 AAGTCACAGAGCTAAGCTGCTGAGACAGACAAATCCATGGAGCATAT 546
151 AlalysSerLysSerHisValIleIleSerLeuLysThrValLysGly 167
547 GCAAAACATTAGCCATGATGATCATAGACTCAAAACAGCCAAAGGAC 596
167 rlySGluLeuLysLeuAspSerSerLysThrAspAlaLeuIleLysGlu 184
597 CAACAGGTGAAAGCTGAGCCCAAGCATTTTAAAGTTTGCAAGAAAGAC 646
184 ySValAlaValGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
647 GAGTAAAGCTGAGATGTGATTACATTAAGCAACAGTGGGCGCGT 696
201 LysArgValGlyArgCysAspSerPheAlaThrGluTyrAspLeuGluAl 217
697 AAGAGCCAGGAGGTGATGATGATGATGATGATGATGATGATGATGAT 746
217 agGluTyrValProIleProLysGlyGluValHisLysLysGluI 234
747 TGAAGGTATGTCCTTCCCAAAAGGATGTGCAAAAGAAAGAAAGAA 796
234 lValGluAspValThrLeuHisAspLeuAspAlaAlaAsnLacIPro 250
797 TCATCCAGATGTGACCTTGATGATGATGATGATGATGATGATGATG 846
251 GluGlyGlyGluAspIleLeuSerLeuMetGlyGluMetLysProAr 267
847 CAGGGGGAGAAAGATATCTGTCATGATGAGGCGACCTATATAGCCAA 896
267 gLysThrGluIleThrGluLysLeuArgGluLeuValLeuValVal 284
897 GAAGACAGAAATCAGACAGAACTTGAGGGGAGATTATATAGGTGTA 946
284 snArgTyrIleAspGluGlyIleAlaGluLeuValProGlyValLeuPhe 300
947 ACAAGTACATCGACAGGCGATGCTGAGCTGCTCCGGGTGCTGCTT 996
301 lLasGluValHisMetLeuAspIleGluCysPheSerTyrLeuAsnAr 317
997 GTTGAATGAGTCCACATGCTGACATTTGAGTGTACTACTGTCAGCG 1046
317 gAlaLeuGluSerProLeuSerProIleValIleLeuValThrAsnArg 334
1047 CGGCTGAGATCTTTCATGCTCCATGCTCATCTTTCATCCAAACGAG 1096
334 lylLeCysAsnValArgGlyThr..AspMetThrSerProHisGlyIle 349
1097 GCAACTGTGTATCAGAGGACATGAGACATCAATCCCTCAGGCGATC 1146
350 ProValAspLeuLeuAspArgLeuValIleLeuArgThrGluThrTyrG 366
1147 CCTCTTACCTCTGAGACGAGTATGATATATCCGACCATGCTGTATAC 1196
366 yProThrGluMetIleGluIleLeuValAlaLeuArgAlaGluGluI 383

```

```

1197 TCACAGAGAAATGAAACAGATCATTAATAATCCGTGCCACAGGAGAA 1246
383 lAspMetAspGluGluSerLeuAlaTyrLeuGlyValIleGlyGluIn 399
1247 TCACATCAGTGAAGGAGGACCTAGACCACTGGGAGATGGACCAAG 1296
400 ThrSerLeuArgHisAlaIleGluLeuIleSerProAlaSerValAla 416
1297 ACCACACTGAGTACTAGTACGACGCTGACCCCGCCCAACTGCTGC 1346
416 rlyThrAsnGlyArgGluLysIleCysLysAlaAspLeuGluValS 433
1347 TAAATCAACGGAGAGACAGCATTTGAAAGAGCATGTGAGAGATCA 1396
433 ergLeuTyrLeuAspAlaLysSerSerLarArgLeuGluGluIn 449
1397 GTGACCTTTCTATGATGCCAAGTCTCCGCAAAATCTGCTGACCA 1446
450 GluGluArgTyrIle 454
1447 CAGGATTAAGTACATG 1461
seq_name: gb_pr7:BC000519
seq_documentation_block:
LOCUS BC000519 2503 bp mRNA PRI 16-MAR-2001
DEFINITION Homo sapiens, RUvB (E coli homolog)-like 1, clone MGC:8557, mRNA, complete cds.
ACCESSION BC000519
VERSION BC000519.1 GI:12653494
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK
COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue procurement: DCRD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
nisc_mgc@nih.gov
Contact:
Shevchenko,Y., Wehlerdy,K.D., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
Lim,M., Maduro,Q.L., Mastaglio,C., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Snyder,B., Stantrifop,S., Thomas,P.J.,
Tionson,E.E., Touchman,J.W., Tsugeon,C., Vogt,J.L., Walker,M.A.,
Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 1 Row: j Column: 20.
location/Qualifiers
1..2503
/organism="Homo sapiens"
/db_xref="locusID:8607"
/db_xref="taxon:9606"
/clone="MGC:8557 IMAGE:2822889"
/tissue_type="Lung, small cell carcinoma"
/clone_id="NIH_MGC_7"

```

FEATURES

SOURCE

/lab_host="DH10B-R"
/note="Vector: pOT87"
81..1451
CDS

/codon_start=1
/product="RuVb (E coli homolog)-like 1"
/protein_id="AAH00519.1"
/db_xref="GI:12653495"

/translation="MKIEYKSTTKTORTASHSHVKGILDESLAKOASGLVGOEN
AREACGYVELIKSKMAGRAVLALAPPGKTAALALAOELGSKVPPCWGSEVY
STELKKREVMENPRRAIGRIKETKEVKEVEYTELTPEETENPMGCGYKTTLSHTIG
LTKAKGTQLKLPDSIPESLQKERVAGDVIYEANSVAVRQGRCDIYATFEDLEAE
EYVLPKGVGDKHKKETIIOVTLHDLVANARPGODILSMGOLMKPKETETDLRAE
GEINKVYVKYIDOGIAELVGVLFVEVHMLDECTYLRHLESSIAPIVIFASNRG
NCVIRGHEDTTSPHGILPDLIDRVMTIRFMILYPOEMKOIKIRAOEGINISEEALN
HGEIGKTTLRYSVOLTPANLAKINKDSEIERHVEISELFTDANSKAILADQ
QDKYMK"

BASE COUNT 681 a 642 c 644 g 536 t
ORIGIN

alignment_scores:
Quality: 1723.50 Length: 455
Ratio: 4.113 Gaps: 1
Percent Similarity: 92.088 Percent Identity: 72.527

alignment_block:
us-09-589-510-4 x BC000519 ..

Align seg 1/1 to: BC000519 from: 1 to: 2503

```

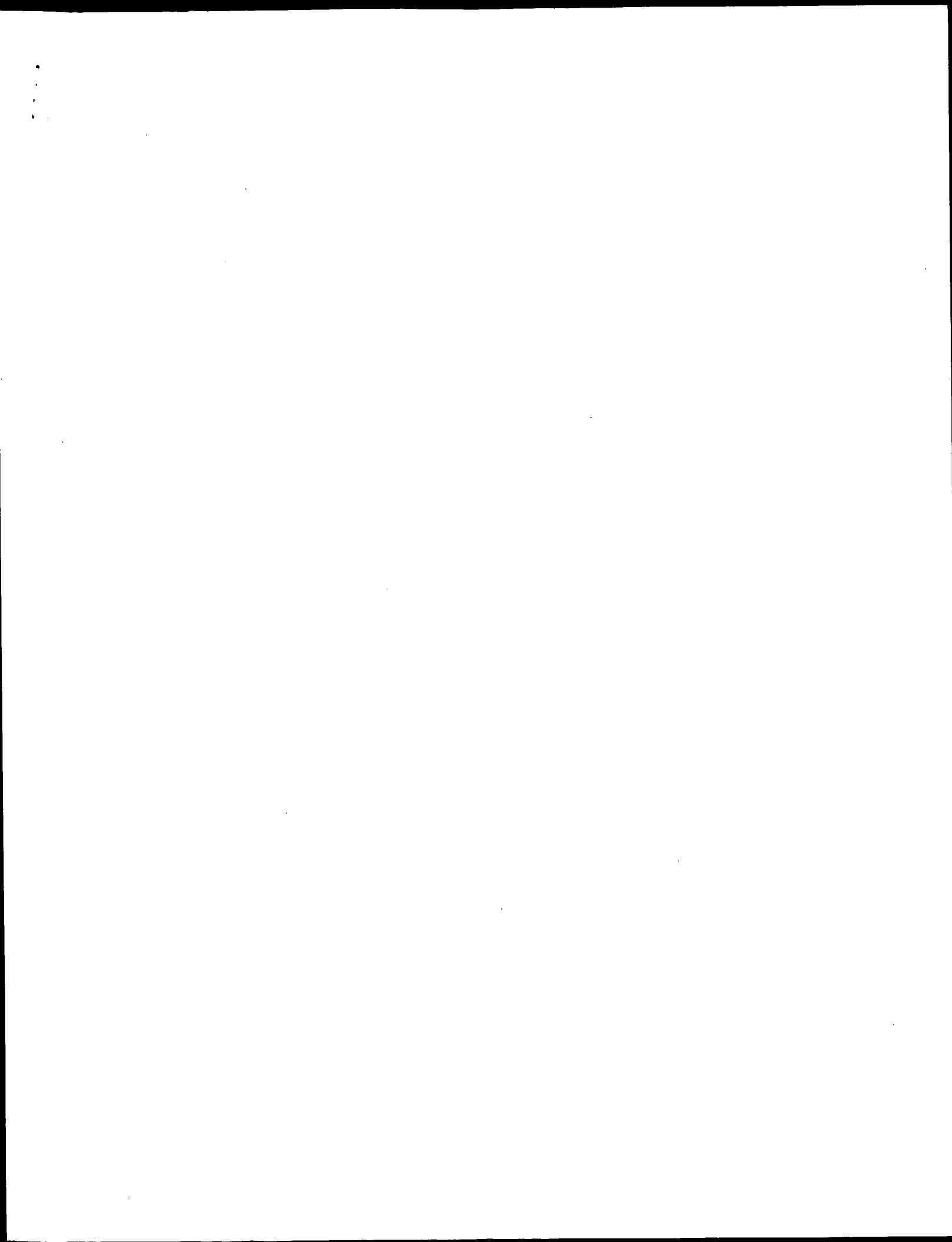
1 MetATGllleGluGluValGlnSerThrSerLysLysGlnArgIleAlaThr 17
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 ATGAAGATTGAGAGAGTGAAGACGACACGACGACGACGACGACGACGACG 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 RHISRHISLIElySGlyLeuGlyLeuAspAlaAsnGlyMetAlaIleA 34
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
131 CCACAGCCAGCGAAAGGGCTGGGCTGGACGAGAGCGGCTGGCCCAAGC 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 lAlaValAlaGlyPheValGlyGlnAlaAlaAlaArgGluAlaAlaGly 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 AGCGGCGCTCAGGCGTCTGGGCCAGAGAGACGCGCGAGGCGATGTGGC 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 LeuAlaValAspMetIleArgGlnLysLysMetAlaGlyArgAlaVal 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
231 GTCAATAGTAGAATTAATCAAAAGCAAGAAATGGCTGGAGAGCTGTCTT 280
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 uLeuAlaGlyProProAlaThrGlyLysThrAlaLeuAlaLeuGlyIleA 84
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
281 GTTGGAGGACCTCTCTGAACTGGCAAGACGCTCTGCTGTGGCTATTTG 330
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 lAGlnGluLeuGlySerLysValProPheCysPrometValGlySerGlu 100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 CTCAGGAGCTGGTAGTAGTCCCTTCTGCCAATGGTGGGAGTGA 380
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 ValIlySerSerGluValLysLysThrGluValLeuMetGluAsnPheAr 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
381 GTTACTCACTGAGATCAAGAAAGACAGAGGTGCTGATGAGAACTCCG 430
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 gArGAlAlleGlyLeuArgIleLysGluAsnLysGluValTyrGluGly 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
431 CAGGCGCATTTGGCTGGAAATAAAGAGACCAAGAAAGTTATCAAGGTG 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 lValAlThrGluLeuSerProGluGluAlaGlnSerThrThrGlyGlyTyr 150
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
481 AAGTCACAGAGCTAACTCCGTGTAGACAGAGAAATCCCATGGGAGATAT 530
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 AlAlaSerIleSerHisValIleIleSerLeuLysThrValLysGlyThr 167
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
531 GGCAAAACCATTAACCATGTGATCATAGACTCAAAACAGCAAGAGAAC 580
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 rLySGlnLeuLysLeuAspSerSerIleTyrAspAlaLeuIleLysGlu 184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
581 CAACAGTTGAACCTGGACCCGACATTTTGAAGTTTGCAGAAAGAGC 630

```

```

184 ySValAlaValAlaGlyAspValIleTyrIleGluAlaAsnSerGlyAlaVal 200
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
631 GAGTAAAGACTGGAGATGTGATTACATTGAAAGCAACAGATGGGCGCTGT 680
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 LysArgValAlaGlyArgCysAspSerPheAlaThrGluTyrAspLeuGluVal 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
681 AAGAGGACAGGCGCAGGTGTGATCTATGCCACAGAAATTCGACCTTGAGC 730
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 aGluGluTyrValProIleProLysGlyGluValHisLysLysGluIle 234
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
731 TGAAGACTATGTCCCTTGGCAAAAGGAGATGTGGCAACAAAGAAAGAA 780
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
234 lValAlaGlnAspValThrLeuHisAspLeuAspAlaAlaAsnAlaGlnPro 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
781 TCACCAAGATGTGACCTTCATGATCTTGATGTGGCTAATGGCGGCC 830
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 GlnGlyGlyGlnAspIleLeuSerLeuMetGlyGlnMetLysProAr 267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
831 CAGGCGGACAGATATCTCTGTCATGATGGCCAGCTTAATGAAGCCAA 880
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
267 gLysThrGluIleThrGluLysLeuArgGlnGluIleAsnLysValAla 284
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
881 GAAGACAGAAATCAACAGCAAACTTCGAGGGGAGATTAAATAGGTGTGA 930
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
284 snArgTyrIleAspGluGlyIleAlaGluLeuValProGlyValLeuPhe 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
931 ACAAGTACATGACAGCGGCGATTGCTGTGGTGTCCGGGTGTGTGTT 980
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 lLeaspGluValHisMetLeuAspIleGluCysPheSerTyrLeuAsnAr 317
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
981 GTTGATGAGGTCCACATCTCGACATTTGACTGCTTACCTGACTGACCG 1030
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
317 gAlaLeuGluSerProLeuSerProIleValIleLeuAlaThrAsnArg 334
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1031 CGCCCTGAGACTTCTTCACTGCTCCATCGTCATCTTGATCAACCGAG 1080
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
334 lYlLeCysAsnValArgGlyThr...AspMetThrSerProHisGlyIle 349
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1081 GCAACTGTGTATCAAGAGGACGACGTGAGACATCAATCCCTCACGGCATC 1130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
350 ProValAspLeuLeuAspArgLeuValIleIleArgThrGluThrTyrGl 366
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1131 CCCTTGACCTTCTGACCGAGTATGATATCCGACCATGCTGTATAC 1180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
366 yProThrGluMetIleGlnIleLeuAlaIleArgAlaGlnValGluGlu 383
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1181 TCCACAGAAATGMAACAGATCATTAATAATCCGTCGCCAGAGGAGAA 1230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
383 lAspMetAspGluGlnSerLeuAlaTyrLeuGlyGluIleGlyGlnGln 399
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1231 TCACATCTGATGTGAGGAGGACCTGAAACCCTGGGGAGATTTGGCAAC 1280
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
400 ThrSerLeuArgHisAlaIleGlnLeuIleSerProAlaSerValAlaSe 416
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1281 ACCACACTGAGTACATTCAGTGCAGTGTGACCCGCCCACTTGCTTGC 1330
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
416 rLysThrAsnGlyArgGluLysIleCysLysAlaAspLeuGlnGluValS 433
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1331 TAAATAACAAGGGAAGACAGCATTCAGAAAGAGCATGTGCAAGAGATCA 1380
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
433 eArgLysTyrLeuAspAlaLysSerSerAlaArgLeuLeuGlnGln 449
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1381 GTGAACTTTTCTATGATGCCAAGTCTCCGCCAATAATCTGTGGTGACG 1430
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
450 GlnGluArgTyrIle 454
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1431 CAGGATTAACATCATG 1445
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
seq_name: gb_rol:AB001581
seq_documentation_block:
LOCUS AB001581 1567 bp mRNA ROD 24-MAR-1999

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2001, 10:42:30 ; Search time 143.43 Seconds
(without alignments)
8370.270 Million cell updates/sec

Title: US-09-589-510-3

Perfect score: 1912

Sequence: 1 acccaagcgtccgcaattt.....aaaaaaaaaaaaaaaaaa 1912

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

- 1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910.8	99.9	1912	22	AA02567
2	1686.6	88.2	1845	22	AA02566
3	1630	85.3	1869	22	AA02570
4	1362.2	71.2	1886	22	AA02568
5	1351.8	70.7	1898	22	AA02569
6	783.6	41.0	1612	21	AA038992
7	622.6	32.6	1587	20	AA038992
8	612.4	32.0	1730	20	AA038992
9	612.4	32.0	1750	20	AA038992
10	544.4	28.5	1744	21	AA038992
11	524	27.4	1509	20	AA038992

12	336.8	17.6	1213	21	AA03700	Aspergillus oryzae
13	244.6	12.8	1416	21	AA03017	Yeast HEL50 DNA he
14	228.2	11.9	1492	21	AA03016	Human HEL50 DNA he
15	166	8.7	1280	22	AA074190	DNA encoding evlro
16	157.8	8.3	1167	20	AA084507	Human secreted pro
17	137.2	7.2	335	21	AA080854	Fusarium venenatum
18	129.6	6.8	1002	20	AA030161	Human secreted pro
19	128.4	6.7	429	21	AA030298	Human secreted pro
20	118.8	6.2	433	21	AA030399	Human secreted pro
21	109.6	5.7	749	21	AA074528	Fusarium venenatum
22	108.6	5.7	313	14	AA060576	Human brain expres
23	101.6	5.3	936	22	AA058257	Oligonucleotide D1
24	101.4	5.3	936	22	AA058254	Oligonucleotide D1
25	101.4	5.3	938	22	AA058255	Oligonucleotide D1
26	101.2	5.3	936	22	AA058252	Oligonucleotide D1
27	99.8	5.2	936	22	AA058259	Oligonucleotide D1
28	99.8	5.2	936	22	AA058262	Oligonucleotide D2
29	98.6	5.2	936	22	AA058252	Oligonucleotide D2
30	98.6	5.2	936	22	AA058254	Oligonucleotide D1
31	98.6	5.2	936	22	AA058257	Oligonucleotide D1
32	98.6	5.2	936	22	AA058259	Oligonucleotide D2
33	98.6	5.2	936	22	AA058262	Oligonucleotide D2
34	98.6	5.2	938	22	AA058255	Oligonucleotide D1
35	54.6	2.9	769	21	AA028122	Arabidopsis thalia
36	53	2.8	844	21	AA029112	CDNA encoding C-te
37	51.4	2.7	721	21	AA073912	Human secreted pro
38	51.4	2.7	1696	21	AA016214	Human prostate can
39	51.4	2.7	3891	21	AA076424	Human ORFX ORF1979
40	51.2	2.7	309	21	AA094021	Cat flea hindgut a
41	51.2	2.7	310	21	AA094563	Cat flea hindgut a
42	50.4	2.6	8749	22	AA087129	Nucleotide sequenc
43	50.4	2.6	8941	22	AA087131	Plasmid pGC1037 nu
44	50.4	2.6	9700	22	AA090958	
45	50.4	2.6	9749	22	AA090959	

ALIGNMENTS

RESULT 1	
AA02567	standard; cDNA: 1912 BP.
ID	AA02567
XX	
AC	AA02567;
XX	
DT	02-MAY-2001 (first entry)
XX	
DE	Maize RuVB orthologue #2 cDNA.
XX	
KW	Maize; RuVB orthologue; branch migration; heteroduplex extension;
KW	homologous recombination; transformation; transgenic plant; ss.
OS	Zea mays.
XX	
FH	Key
FT	CDS
FT	Location
FT	94..1461
FT	/*tag= a
FT	/product= "Maize RuVB orthologue protein #2"
XX	
PN	W0200105975-A1.
XX	
PD	25-JAN-2001.
XX	
PF	13-JUN-2000; 2000WO-US16271.
XX	
PR	16-JUL-1999; 9905-0144112.
XX	
PA	(PION-) PIONEER HI-BRED INT INC.
XX	
PI	Mahajan PB;
XX	
DR	WPI: 2001-159537/16.
DR	P-PSDB: AAY72563.

XX Novel maize RuVb nucleic acid useful for modulating levels of maize
 PT RuVb in plants, as probes or amplification primers in the detection,
 PT quantitation or isolation of gene transcripts

PS Claim 1; Page 70-72; 87pp; English.

XX The present sequence is a Zea mays RuVb orthologue #2 cDNA. RuVb along
 CC with RuVb catalyses the branch migration process, also known as
 CC heteroduplex extension, in homologous recombination. RuVb is used for
 CC the control of homologous recombination or transformation efficiency in
 CC transgenic plants. The RuVb nucleotide may be used as probes or
 CC amplification primers for detecting, quantifying or isolating gene
 CC transcripts, in detecting deficiencies in the mRNA level during screening
 CC for desired transgenic plants, for detecting gene mutations, for
 CC monitoring upregulation of expression or changes in enzyme activity, for
 CC detecting any number of allelic variants, orthologues or paralogues of
 CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
 CC also be used for recombinant expression of its encoded polypeptide, or
 CC for use as immunogen in preparing and/or screening of antibodies, and in
 CC sense or antisense suppression of one or more genes in a host cell,
 CC tissue or plant. The RuVb proteins may be used in assays to agonise or
 CC antagonise the enzyme function, or as immunogens or antigens for
 CC screening antibodies.

SQ Sequence 1912 BP: 547 A; 385 C; 515 G; 462 T; 3 other;

Query Match 99.9%; Score 1910.8; DB 22; Length 1912;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1912; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 accacagctccgaataattgttcgagcgagagccgagagagagcaatccacagaa 60
 DB 1 accacagctccgaataattgttcgagcgagagccgagagagagcaatccacagaa 60
 QY 61 acagagagcgataaccgagcggttcgagcgagatcgagagagagcgagcgacac 120
 DB 61 acagagagcgataaccgagcggttcgagcgagatcgagagagagcgagcgacac 120
 QY 121 tcgaagaacgagcgatccgacccacacacacacacacacacacacacacacacac 180
 DB 121 tcgaagaacgagcgatccgacccacacacacacacacacacacacacacacacac 180
 QY 181 gggatgagcgatggttcgagcgagcggttcgagcgagcgagcgagcgagcgagc 240
 DB 181 gggatgagcgatggttcgagcgagcggttcgagcgagcgagcgagcgagcgagc 240
 QY 241 gggatgagcgatggttcgagcgagcggttcgagcgagcgagcgagcgagcgagc 300
 DB 241 gggatgagcgatggttcgagcgagcggttcgagcgagcgagcgagcgagcgagc 300
 QY 301 ggtccgcccgcacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 360
 DB 301 ggtccgcccgcacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 360
 QY 361 aaggtcccttcgtcccatgtagatcagaagtgtatctcctcgtgaggttaagaact 420
 DB 361 aaggtcccttcgtcccatgtagatcagaagtgtatctcctcgtgaggttaagaact 420
 QY 421 gaggtcgtatggaataattccgttagagctataggttcgtatataaagaacaagag 480
 DB 421 gaggtcgtatggaataattccgttagagctataggttcgtatataaagaacaagag 480
 QY 481 gttatgaagagaggttaactgaattcccccagaagaggtgagtaaacactggtgga 540
 DB 481 gttatgaagagaggttaactgaattcccccagaagaggtgagtaaacactggtgga 540
 QY 541 tatgcaaaaagcattagcattatcaatcgaatgaagagcggttaaaagggactaaagaa 600
 DB 541 tatgcaaaaagcattagcattatcaatcgaatgaagagcggttaaaagggactaaagaa 600
 QY 601 ctgaagttagattcttcaattatgactgactgataagaagaagtgagcggtgagat 660

DB 601 ctgaagttagattcttcaattatgactgactgataagaagaagtgagcggtgagat 660
 QY 661 gtatataatcgaagcaaatagtgagcgagtgaaagagtgtagatgtatcttct 720
 DB 661 gtatataatcgaagcaaatagtgagcgagtgaaagagtgtagatgtatcttct 720
 QY 721 gctacaagaatacgttgaagcgtgagagatgttctctatccccaaagtgagtcac 780
 DB 721 gctacaagaatacgttgaagcgtgagagatgttctctatccccaaagtgagtcac 780
 QY 781 aagaaagaagaattgttcgaagatgttcacacttcagacttcagtgagcaaatgtcag 840
 DB 781 aagaaagaagaattgttcgaagatgttcacacttcagacttcagtgagcaaatgtcag 840
 QY 841 ccacaaagtgcccaagatatttgcctctatggcgagagatgagaaacccgaaagact 900
 DB 841 ccacaaagtgcccaagatatttgcctctatggcgagagatgagaaacccgaaagact 900
 QY 901 gaattccgcaaaatactacgccaagaattaaagtggttaaatagatatcatgtgaa 960
 DB 901 gaattccgcaaaatactacgccaagaattaaagtggttaaatagatatcatgtgaa 960
 QY 961 ggaattgcagagcttgtagcttggtttgttcattgtagaggtccacatgttgatc 1020
 DB 961 ggaattgcagagcttgtagcttggtttgttcattgtagaggtccacatgttgatc 1020
 QY 1021 gaattgtttcttattcttaccggtcatgtagagcccatcacaatcgttactt 1080
 DB 1021 gaattgtttcttattcttaccggtcatgtagagcccatcacaatcgttactt 1080
 QY 1081 gctacaataagggaaatgtaatagttaagaagaactgatatgacaagttccacatgtata 1140
 DB 1081 gctacaataagggaaatgtaatagttaagaagaactgatatgacaagttccacatgtata 1140
 QY 1141 ccggtgagcttcttaataatgctgtgtgattatcgagacgaagacatatggcctactag 1200
 DB 1141 ccggtgagcttcttaataatgctgtgtgattatcgagacgaagacatatggcctactag 1200
 QY 1201 atgatacagatattgtctatccgagcacaaagtggagagatgatatgtagaagaagt 1260
 DB 1201 atgatacagatattgtctatccgagcacaaagtggagagatgatatgtagaagaagt 1260
 QY 1261 ctgtctatttgaagcgagatcgagacagacacatctttagacatgctatcaattgata 1320
 DB 1261 ctgtctatttgaagcgagatcgagacagacacatctttagacatgctatcaattgata 1320
 QY 1321 tcacctgccaagcggtgttccaagaactaatggaagagaataatctgcaaggtgacttc 1380
 DB 1321 tcacctgccaagcggtgttccaagaactaatggaagagaataatctgcaaggtgacttc 1380
 QY 1381 gaggaagtcagtggtctatttgaatgccaatctcgtgctcgtgctcgaagagcaa 1440
 DB 1381 gaggaagtcagtggtctatttgaatgccaatctcgtgctcgtgctcgaagagcaa 1440
 QY 1441 caagaaagataatcaactagatttgaatcctctgtcgtggaagttctgaagaagtata 1500
 DB 1441 caagaaagataatcaactagatttgaatcctctgtcgtggaagttctgaagaagtata 1500
 QY 1501 gttgcaagtcgaagatcatcagtagtgcattgtagcttcacaaaggttcaatgactcag 1560
 DB 1501 gttgcaagtcgaagatcatcagtagtgcattgtagcttcacaaaggttcaatgactcag 1560
 QY 1561 tcttgagcgagacatttcgggggaaacggttgaaatttgcaatgctgctgtgtgta 1620
 DB 1561 tcttgagcgagacatttcgggggaaacggttgaaatttgcaatgctgctgtgtgta 1620
 QY 1621 gttcccaagaagaacttggttccgcatatgtctgttaacgacagcttcgctgattaga 1680
 DB 1621 gttcccaagaagaacttggttccgcatatgtctgtgttaacgacagcttcgctgattaga 1680
 QY 1681 ttggtcacggtgcagaagatgctgctgtgtgtttatcttcgtacatcggttccgagct 1740

Db 1681 ttgtaccggttcgacgaatcgctgtgtgtttatcttgcacatcggtgcgagct 1740
Oy 1741 ctgctccacgggtgtgtatctggccgaacctatcttgttaacatgataatgataag 1800
Db 1741 ctgctccacgggtgtgtatctggccgaacctatcttgttaacatgataatgataag 1800
Oy 1801 attcttaagaatgaacttcgacttcgttatttcttaaatgttcataagaataag 1860
Db 1801 attcttaagaatgaacttcgacttcgttatttcttaaatgttcataagaataag 1860
Oy 1861 aatgtttctcaacmctwtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912
Db 1861 aatgtttctcaacmctwtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1912
RESULT 2
AAD02566
ID AAD02566 standard; cDNA: 1845 BP.
AC AAD02566;
XX 02-MAY-2001 (first entry)
D7 Maize RuVB orthologue #1 cDNA.
DE Maize; RuVB orthologue; branch migration; heteroduplex extension;
XX homologous recombination; transformatio; transgenic plant; ss.
KW
OS Zea mays.
XX
FH Key Location/Qualifiers
FT 85..1452
FT CDS /*tag= a
FT /product= "Maize RuVB orthologue protein #1"
XX
PN W0200105975-A1.
XX 25-JAN-2001.
PD
PF 13-JUN-2000; 2000WO-US16271.
XX
PR 16-JUL-1999; 99US-0144112.
XX
PA (PION-) PIONEER HI-BRED INT INC.
PI Mahajan PB;
PI
PI WPI: 2001-159537/16.
DR P-PSDB; AAY72562.
DR
PT Novel maize RuVB nucleic acid useful for modulating levels of maize
PT RuVB in plants, as probes or amplification primers in the detection,
PT quantitation or isolation of gene transcripts -
PS
PS Claim 1; Page 67-69; 87pp; English.
XX
XX The present sequence is a Zea mays RuVB orthologue #1 cDNA. RuVB along
CC with RuVA catalyses the branch migration process, also known as
CC heteroduplex extension, in homologous recombination. RuVB is used for
CC the control of homologous recombination or transformation efficiency in
CC transgenic plants. The RuVB nucleotide may be used as probes or
CC amplification primers for detecting, quantifying or isolating gene
CC transcripts, in detecting deficiencies in the mRNA level during screening
CC for desired transgenic plants, for detecting gene mutations, for
CC monitoring upregulation of expression or changes in enzyme activity, for
CC detecting any number of allelic variants, orthologues or paralogues of
CC the gene, or for site directed mutagenesis in eukaryotic cells. It may
CC also be used for recombinant expression of its encoded polypeptide, or
CC for use as immunogen in preparing and/or screening of antibodies, and in
CC sense or antisense suppression of one or more genes in a host cell,
CC tissue or plant. The RuVB proteins may be used in assays to agonise or
CC antagonise the enzyme function, or as immunogens or antigens for
CC screening antibodies.

XX
SQ Sequence 1845 BP; 530 A; 378 C; 490 G; 445 T; 2 other;
Query Match 88.2%; Score 1686.6; DB 22; Length 1845;
Best Local Similarity 96.5%; Pred. No. 0;
Matches 1758; Conservative 3; Mismatches 42; Indels 18; Gaps 3;
Oy 46 ggcagctccacagaaacagagagcgataacccggcgcgctgtggcgagatcgag 105
Db 37 ggcagctccacagaaacagagagcgataacccggcgctgtgtggcgagatcgag 96
Oy 106 gagggtcagtcgaccccgaaagagcgatccgaccccaaccacatcaaggagctc 165
Db 97 gagggtcagtcgaccccgaaagagcgatccgaccccaaccacatcaaggagctc 156
Oy 166 ggcctcagccacatgagatgagatctgtgctggcgcggtgtgtggcgagcgcg 225
Db 157 ggcctcagccacatgagatgagatctgtgctggcgcggtgtgtggcgagcgcg 216
Oy 226 ggcgcgagcgcgcgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 285
Db 217 ggcgcgagcgcgcgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 276
Oy 286 ggcgtgtcctccttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 345
Db 277 ggcgtgtcctccttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 336
Oy 346 caggagctcgcgcagcagagctccttctgtctatgtatgatacgaagtgatcctcg 405
Db 337 caggagctcgcgcagcagagctccttctgtctatgtatgatacgaagtgatcctcg 396
Oy 406 gagggtcaagaacacgtgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 465
Db 397 gagggtcaagaacacgtgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 456
Oy 466 aaggaaacacacagaggtttatgaagagaggttactgaacttcccgagagagctag 525
Db 457 aaggaaacacacagaggtttatgaagagaggttactgaacttcccgagagagctag 516
Oy 526 agtacaactgtgtgatalatgcaaaaagcattagcattgaatcatcagcttaagactgt 585
Db 517 agtacaactgtgtgatalatgcaaaaagcattagcattgaatcatcagcttaagactgt 576
Oy 586 aaggagactgaagcaactgaagttgaattcttcaatttatgtatgtctgtatcaaggaag 645
Db 577 aaggagactgaagcaactgaagttgaattcttcaatttatgtatgtctgtatcaaggaag 636
Oy 646 gtgcagtgaggtgagttatatatactgaagcaaatagtgagcaggtgaagagttgt 705
Db 637 gtgcagtgaggtgagttatatatactgaagcaaatagtgagcaggtgaagagttgt 696
Oy 706 agatgtatctcttctgtacagaaatcagatcttgaagcttgaaggtatgtctccatccc 765
Db 697 agatgtatctcttctgtacagaaatcagatcttgaagcttgaaggtatgtctccatccc 756
Oy 766 aagggtgaggttccataaagaaaaaagaaattgttgcagaggtgtcacaactcatgacctgtat 825
Db 757 aagggtgaggttccataaagaaaaaagaaattgttgcagaggtgttgcacaactcatgacctgtat 816
Oy 826 gcaagcaaatgtctcaacccaaggtgtgccaagatatgtgtccctatagggcgcaagatgt 885
Db 817 gcaagcaaatgtctcaacccaaggtgtgccaagatatgtgtccctatagggcgcaagatgt 876
Oy 886 aaaccacgaagaactgaatcaccgaaaaactacgccaagaagaataaaggtgttaaat 945
Db 877 aaaccacgaagaactgaatcaccgaaaaactacgccaagaagaataaaggtgttaaat 936
Oy 946 agatatacgtatgaagaagatgcagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1005
Db 937 agatatacgtatgaagaagatgcagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 996
Oy 1006 cacatgttgatatacgaatgtttcttatcttaacgcgtgatgtgagagccatataca 1065

QY 346 caggagctcgccagcaaggctcccttctgtccatgtagatcagaagtgtactctcg 405
 |||||||
 Db 265 caggagctcgccagcaaggctcccttctgtccatgtagatcagaagtgtactctcg 324
 QY 406 gaggtaaaagaaactgaagtgctgtatgtaaaattccgtagagctataagtttgcgtata 465
 |||||||
 Db 325 gaggtaaaagaaactgaagtgctgtatgtaaaattccgtagagctataagtttgcgtata 384
 QY 466 aaggaaacaaagagtttataagagagaggtgtactgaaattcccccgaagaagctgag 525
 |||||||
 Db 385 aaggaaacaaagagtttataagagagaggtgtactgaaattcccccgaagaagctgag 444
 QY 526 agtacaaactgtgtatgataatgcaaaagcattagccatgtatcatcagcttaagagctgt 585
 |||||||
 Db 445 agtacaaactgtgtatgataatgcaaaagcattagccatgtatcatcagcttaagagctgt 504
 QY 586 aaggagacaaagcaactgaagttaattcttcaattatgataagctgtatcagaagaaag 645
 |||||||
 Db 505 aaggagacaaagcaactgaagttaattcttcaattatgataagctgtatcagaagaaag 564
 QY 646 gtgagcagtggtgtatgataatgataatgcaaaagcattagccatgtatcatcagcttaag 705
 |||||||
 Db 565 gtgagcagtggtgtatgataatgataatgcaaaagcattagccatgtatcatcagcttaag 624
 QY 706 agatgtgattcttctgtctacagaatacgaacttgaaagctgaagagatgttctctatcccc 765
 |||||||
 Db 625 agatgtgattcttctgtctacagaatacgaacttgaaagctgaagagatgttctctatcccc 684
 QY 766 aagggtgaagtcacataagaagaaagaaattgtgtcagagatgtgtccacactgtacctgtat 825
 |||||||
 Db 685 aagggtgaagtcacataagaagaaagaaattgtgtcagagatgtgtccacactgtacctgtat 744
 QY 826 gcaagcaaatgtctcagcccaaggtgtgtccagaagattttgtccctatgtgcagatgtatg 885
 |||||||
 Db 745 gcaagcaaatgtctcagcccaaggtgtgtccagaagattttgtccctatgtgcagatgtatg 804
 QY 886 aaacccagaaagactgaaataccgcgaagaaactgaagccagaagaataaaggtgtataat 945
 |||||||
 Db 805 aaacccagaaagactgaaataccgcgaagaaactgaagccagaagaataaaggtgtataat 864
 QY 946 agatatcagatgaaggaattgtcagagctgttactgtgtgtttgttcattgtatgaaggtc 1005
 |||||||
 Db 865 agatatcagatgaaggaattgtcagagctgttactgtgtgtttgttcattgtatgaaggtc 924
 QY 1006 cacatgtgtgatalcogaatgttttcttcatatcaccgtgcatgtgagagcccatatca 1065
 |||||||
 Db 925 cacatgtgtgatalcogaatgttttcttcatatcaccgtgcatgtgagagcccatatca 984
 QY 1066 ccaatcgtgatactgtctacaaataaggggaatatagttaagaggaactgtatagaca 1125
 |||||||
 Db 985 ccaatcgtgatactgtctacaaataaggggaatatagttaagaggaactgtatagaca 1044
 QY 1126 agtcacaatgtataccggtgtgagatcttctagataggtgtgtattatcgacaagagaca 1185
 |||||||
 Db 1045 agtcacaatgtataccggtgtgagatcttctagataggtgtgtattatcgacaagagaca 1104
 QY 1186 tatggccctactgagatgatalcagataatgtgtatccgagacacaagtgagggagattgat 1245
 |||||||
 Db 1105 tatggccctactgagatgatalcagataatgtgtatccgagacacaagtgagggagattgat 1164
 QY 1246 atggatgaagaagaagcttctgtattataggcgagatcgagacagacacatcttggagacat 1305
 |||||||
 Db 1165 atggatgaagaagaagcttctgtattataggcgagatcgagacagacacatcttggagacat 1224
 QY 1306 gctatcaatgtatatacactgtgcagcggtgtgtcacaagcctaataagagagagaaatc 1365
 |||||||
 Db 1225 gctatcaatgtatatacactgtgcagcggtgtgtcacaagcctaataagagagagaaatc 1284
 QY 1366 ttgcaaggtgtatctcgaggaagtcagtggtctatatttgatgtgccaatccctcgctcg 1425
 |||||||
 Db 1285 ttgcaaggtgtatctcgaggaagtcagtggtctatatttgatgtgccaatccctcgctcg 1344
 QY 1426 ctgtccagagagacaagaagaatacatalcactagatttgatctctctgtgtgaagt 1485

|||||||
 Db 1345 ctgtcccgagggacaacaagaagatacacccttagatttggctacactgtctggaagt 1404
 QY 1486 ctgaaaggaatgtagttgtccagctcgaaagctcatctagttgcatgtatctgtccacag 1545
 |||||||
 Db 1405 ctgaaaggaatgtagttgtccagctcgaaagctcatctagttgcatgtatctgtccaca 1462
 QY 1546 ttcatagctactgtgtcttggagcgagacacatttcggggggaacggttgaatttgcag 1605
 |||||||
 Db 1463 -----gtcttggagcgagacacatttcggggggaacggttgaatttgcag 1509
 QY 1606 tgcctgtgtgttagttagctccaragaagactgtgtcccgcatattgtctgtcagcac 1665
 |||||||
 Db 1510 tgcctgtgtgttagttagctccaragaagactgtgtcccgcatattgtctgtcagcac 1569
 QY 1666 tgttcgctgattagattgtgtaccggtgcaggaatttcgctgtgtg- ttttatactgtc 1724
 |||||||
 Db 1570 tgttcgctgattagattgtgtaccggtgcaggaatttcgctgtgtg- ttttatactgtc 1629
 QY 1725 catcggtgtctcggaatc--tgcttccaggggtgtgtattgtgcccgaacctatcttgtaa 1782
 |||||||
 Db 1630 catcggtgtctcggaatcgtgtccctccacgggtgtgtattgtgcccgaacctatcttgtaa 1689
 QY 1783 ccattgataatgataaggaattcttaccagaatgtcaactgtgcatgtgattatattcctaaa 1842
 |||||||
 Db 1690 ccattgataatgataaggaattcttaccagaatgtcaactgtgcatgtgattatattcctaaa 1749
 QY 1843 tgtccataaagcaatacgaagatttcttacaacmtwtaa 1882
 |||||||
 Db 1750 tgtccataaagcttaacaagaatttcttacaacatataga 1789

RESULT 4
 AAD02368
 ID AAD02568 standard; cDNA; 1886 BP.
 XX
 AC AAD02568;
 XX
 DT 02-MAY-2001 (first entry)
 XX
 DE Maize RuVB orthologue #3 cDNA.
 XX
 KW Maize: RuVB orthologue; branch migration; heteroduplex extension;
 KW homologous recombination; transformation; transgenic plant; ss.
 XX
 OS Zea mays.
 XX
 FH Key Location/Qualifiers
 FT CDS 82..1449
 FT /*tag= a
 FT /product= "Maize RuVB orthologue protein #3"
 PN W0200105975-A1.
 PD 25-JAN-2001.
 XX
 PF 13-JUN-2000; 2000MO-US16271.
 XX
 PR 16-JUL-1999; 9905-0144112.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Mahajan PB;
 XX
 DR WPI: 2001-159537/16.
 DR P-PSDB: AAY72564.
 XX
 PT Novel maize RuVB nucleic acid useful for modulating levels of maize
 PT RuVB in plants, as probes or amplification primers in the detection,
 PT quantitation or isolation of gene transcripts -
 XX
 PS Disclosure: Page 74-76; 87pp; English.
 XX

QY	682	atgtgagcagctgaaaagagtttgtagatgtgattcttcttgctcaacgaatacagatctgaa	741
Db	721	agttagactgtccaacgggtaggttagaagttgatctgtcttcgcacagaaatttgatctggaa	780
QY	742	gtctaaagatgtttccctatccccaaggctgaagtcataaagaanaaagaattgttgag	801
Db	781	gcagaaagaatattgtccacttccccaaggaggtgccaaaagaagaagatagttgag	840
QY	802	gattgcaccttaatgacctttgatctgatcagcaaatgtctcagcccaaggtgtgccaaagattt	861
Db	841	gattccacactccaagaatctcgatgtgaagcaaatgtctcgaccctcaaggtgtgccaaagattt	900
QY	862	ttgtcccttatgtgccagatgatgtgaacacacgaagaagactgtaataccgaaaatactacgc	921
Db	901	cttctctttagtggccaaatgatgaacccggtgaagactgagatcaactcagataagcttcgcg	960
QY	922	caagaataataaagtgtgtaaatagatatatcgaatgaagaattgcagagctgtgacct	981
Db	961	caagaataataaaggtgtgtgaaccgcatatataagaagaagtgtgtcgagagctgtgtcca	1020
QY	982	gtgtttgttctatgtagttaggtgtccacatgttgtagatcgaaatgtttttcttacttaac	1041
Db	1021	ggagttcattatcatctgacgtgaggttccatagatcctgtgaatgtgtcttccataacttgac	1080
QY	1042	cgctgcatgtgagagcccatatataccaatcgttatactgtgtcacaagaatgagggaatattgt	1101
Db	1081	cgctgctcttgagagctcattatctccgatagatgatatattgcacaagaatagaggttttgc	1140
QY	1102	aattgaagaggaactgatatagtacaagtccaatcgaatcgtatccggtgtgattcttbaatg	1161
Db	1141	aacgtaaagaggaactgtatgtaccagcccccacatgtagtccctattgtatcttataagtcga	1200
QY	1162	ctgtgtgattatctggaagaagacatagtgcccctacgtgaatgatabacagatatgtgtac	1221
Db	1201	tgtgttatcatcgcggaactcaatcatatgatccctctcgaaatgatccagattatagccatt	1260
QY	1222	cgagacacaaagtgtgaggaattgtatagatgaatgaagaagctctgtctattttagcgagatc	1281
Db	1261	cgctgcacagtgtgaagaattaacccgtgtatgaagaagtgtgtgttctacttttggtgaagatt	1320
QY	1282	ggacacgcagacatcttttgagacatgtcatattgatatcaccttcgcagcggtgtgtcca	1341
Db	1321	gggcacaagaacttccactaaaggcacgcctgtgacgctctctctccgcagcatgttagcg	1380
QY	1342	aagactaaatggaagaagaaatctgcgaagctcgatctcgaggaagtgcagtgtgctctat	1401
Db	1381	aaatagatggcgtgtgacaarattttgcaagctcgatatagaagaagtaaacatcacctaac	1440
QY	1402	ttgtatgtcacaatctcggctcggctcgtctccagtgagcaaaaagaagaataacacacct	1459
Db	1441	ttgagtgtcataatctctcagcaaaagcttttcagtgaagcaacagaaaaatacatctcat	1498
RESULT 7			
AAAX18198 standard; cDNA to mRNA; 1587 BP.			
AAAX18198:			
10-MAY-1999 (first entry)			
TIP49 coding sequence.			
TIP49; TATA binding protein; TBP binding protein;			
DNA transcription control; ss.			
Rattus sp.			
W099004.19-A1.			
07-JAN-1999.			
25-JUN-1998; 98WO-JP02836.			

Db 691 gtcccttgcacaaggagatgtcacaagaagaagaatcctacagatgtgaccttg 750
 QY 814 catgaccttgatgcacaatgtctcagccaagaagtgcgcaagatatltgtccctatg 873
 Db 751 cagagatgtacgtggtccatgtcgccctcaggtgtggaagatatltgtctatgatg 810
 QY 874 ggcagatgtatgaacacagaagaagactgaatcagcaaaatactagccaagaataat 933
 Db 811 ggcagcttgatgaagcaaaaagaagacagatcacaataactcagaggagatcagac 870
 QY 934 aaggtgttaatatagatatcagtagaagaattgcagagctgtactcgtgtttgttc 993
 Db 871 aaggtgtgaacaataacatcagaccaggtgttcagagctgtccctgtgagtgctctt 930
 QY 994 attgagaggttcacacatgtttgatcagatcgaatttttctacttaaccgtgcatagg 1053
 Db 931 gttagaagatgcacacatgctgatatcagagtgcttaacctactgcagaccgacctgag 990
 QY 1054 agcccatatcaccaatcgtgtactgtctacaataagggagatatgtatgtaagaga 1113
 Db 991 tctccatcgcccccattgtcatcttgcataccagaggaactgtgtcatcaggggc 1050
 QY 1114 ac---tgatatgacaagtccacatgttatcccggtggtatctctagataggtgtgatt 1170
 Db 1051 accgaagacatctactctccacacggtcatcccggtgtgacctgtgaccggtgtgattc 1110
 QY 1171 attcgcagagacatcgtccctactgagatgacagatatgttgctatccgagacaa 1230
 Db 1111 atcagagacatgctgtatcagccacagagatgtagagcagatcattcaagatccgaccag 1170
 QY 1231 gtgagagagattgatatgtagtagaagaagctgtcttatttagcgagatcggacaagag 1290
 Db 1171 acggaagagcatcaatcagtagtagagggccctaaaccactcgaggagattggaccag 1230
 QY 1291 acatcttgagacatgctcattcaattgatatcactcgtccagcggtgtcacaagactaat 1350
 Db 1231 accagcgtgaggtattctcagtagcagctgcagccctgcacacctgtgcgcaagatcac 1290
 QY 1351 ggaagagagaataatcgaaggtctatcgaagagatcgtgagatcgtggtctatttgatcc 1410
 Db 1291 ggggaagagacagattgagaagagagacgtgtgagagatcagcagctcttctatgaagcc 1350
 QY 1411 aatcctcgtgctggtcgtctccagagagacaagaagatatcactagatttgatc 1470
 Db 1351 aagtcctccgcaagatctcgtgcgacagcagagacaagtagatgaagcgtlaagggtt 1410
 QY 1471 t 1471
 Db 1411 t 1411

RESULT 8
 AAX18199 standard; cDNA to mRNA: 1730 BP.
 AC AAX18199;
 DT 10-MAY-1999 (first entry)
 XX TIP49 coding sequence #2.
 DE TIP49
 KM TIP49; TATA binding protein; TBP binding protein;
 KW DNA transcription control; ss.
 OS Rattus sp.
 XX MO9900419-A1.
 PN 07-JAN-1999.
 PD 25-JUN-1998; 98MO-JF02836.
 XX

PR 27-JUN-1997; 97JP-0187398.
 XX
 PA (SOME) SUMITOMO ELECTRIC IND CO.
 XX
 PI Kishimoto T, Makino Y, Niwa S, Tamura T;
 XX
 DR WPI: 1999-095682/08.
 DR P-PSDB: AAW74417.
 XX
 PT TBP-binding protein with DNA helicase and ATPase activities - gene
 PT encoding it, and antibodies recognising it.
 PS
 XX Claim 10; Page 43-45; 64pp; Japanese.
 CC This sequence encodes the TATA-binding protein (TBP) binding protein,
 CC designated TIP49, of the invention. TIP49 and its associated nucleic
 CC acids and antibodies are useful in investigation of the process of DNA
 CC transcription control by TBP in vivo, and in detection of the blocking of
 CC DNA transcription.
 XX
 SO Sequence 1730 BP; 469 A; 401 C; 476 G; 384 T; 0 other;

Query Match 32.0%; Score 612.4; DB 20; Length 1730;
 Best Local Similarity 66.0%; Pred. No. 7.8e-131;
 Matches 902; Conservative 0; Mismatches 461; Indels 3; Gaps 1;

QY 94 atgagatcgaagagagtgtagctgcagctcgaagaagagcagcatcgccaccaaccac 153
 Db 60 atgaagatttgaggggtggaagagacatacgaagcagcagcagctcctccacagccac 119
 QY 154 atcaagagactcggccctccagcccaatgtagatggcattgtcgtgctggtggttcgtg 213
 Db 120 gtgaagaaggctggggggtgagcagagagcggtctgccaagcagcggtcctcagggctgtg 179
 QY 214 ggcacagcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggt 273
 Db 180 ggcacagagagcagcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggt 239
 QY 274 atggcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggt 333
 Db 240 atggcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggt 299
 QY 334 ctggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 393
 Db 300 ctggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 359
 QY 394 gtgtactcctcgtgaggtlcaagaagaactgtgctgtatggaagaatttcgttagatccta 453
 Db 360 gttactcaactgtagatcaagaagacagaggtgtgtatgtagagactcccgagggcatt 419
 QY 454 gtttgcgtataaagaagaacaaagaggtttatgaagaaggttaactgaacttcccca 513
 Db 420 gggctcgtgataaagaagaacaaagaggtttatgaagaaggttaactgaacttcccg 479
 QY 514 gaagaggtcgtgaggtlcaactgtgtatgtatgaagaagcattagccaattatcaccac 573
 Db 480 tgtgagacagagatcccatggtgagatgtgcaaaacattagccaattatcaccac 539
 QY 574 ttaagactgttaagaaggactaagcaactgaagttagaattcttcaattatgatgtcgtg 633
 Db 540 ctcaaaacagccaaaggaacaaacaggttgaactgaagccagcagattttgaagaattg 599
 QY 634 atcaagaagaaggtggtcaggtgtgtatgtatatacatcagaaagcagaatagtggaacagt 693
 Db 600 cagaagaagcaggtgaagaactgtgagatgttatttatttgaagcacaacagtgtggcgtg 659
 QY 694 aaaaagattgtgtagatgttattcttctgtacagatacagatcctgaaagcgtgaagaagt 753
 Db 660 aagagcagggcaggtgtgatacctatgcccagagatcagacttgaagcgtgaagaagt 719
 QY 754 gtctcatcccaaggtggaagtcataagaanaagaagaattgtgcagagatgtcaact 813


```

Db 436 cagattgacgaccgacacagcgagcgtcccaagtgaggcaactgcacccctaagacc 495
Oy 568 atcagcttaagactgtttaaggactaagcaactgaagttatcttcaattatgat 627
Db 496 acagagatgagacacatcagc-----acctggcacaacaagatgatgag 540
Oy 628 gctcgtacacaggaagagtgtagcagtggtgtatgtatatactcgaagcaaatgtaga 687
Db 541 tccctgaccacaggaagagtgtagcagtggtgtatgtatatactcgaagcaaatgtaga 600
Oy 688 gcaatgaaagagtgtagatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 742
Db 601 aagatctccacagctggcgctctctcacaacgcccgcagctacgacgctatggtcc 660
Oy 743 -ctgaagagtggttctcattcccaagagtgtagatccataagaagaaagatgtgacg 801
Db 661 cagaccagagtcgtgtagcagtgtagcagtgtagcagtgtagcagtgtagcagtgtagc 720
Oy 802 gatgcacacttcacgttcgtagcagtgtagcagtgtagcagtgtagcagtgtagcagtg 861
Db 721 accgtgtccctgacagatgcagtgtagcagtgtagcagtgtagcagtgtagcagtgtagc 780
Oy 862 ttgtccctatggcagatgatgaacacacagacagacagacagacagacagacagacagac 921
Db 781 ttctc-----agtgacacagggagtagcagtgtagcagtgtagcagtgtagcagtgtagc 819
Oy 922 caagaataataagtgtagtataatagatataatagatataatagatataatagatataatag 981
Db 820 gagcagatcaatgcaagagtgtagcagtgtagcagtgtagcagtgtagcagtgtagcagtg 879
Oy 982 ggtgtgttgcattgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatg 1041
Db 880 ggaagtgcttcacatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagat 939
Oy 1042 cgtgattgtagagcagcattacacacatgtagatgtagatgtagatgtagatgtagatgtag 1101
Db 940 cgggcctgtagagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtag 999
Oy 1102 aatgtaagaagactgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtag 1161
Db 1000 cgaacccgggacacagctacacagcagcagcagcagcagcagcagcagcagcagcagc 1059
Oy 1162 ctgtgattatcgagacagacatagtgccctactgtagatgtagatgtagatgtagatgtag 1221
Db 1060 ctgtatcgtctccacacacccctacacagcgagaagacagcagcagcagcagcagcagc 1119
Oy 1222 cgagcacaagtgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatg 1281
Db 1120 cgggtcgaggaagagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatg 1179
Oy 1282 gtagcagcagacatctttagacatgtagatgtagatgtagatgtagatgtagatgtagatg 1341
Db 1180 gggctggagagctacgtcgttagcagcagcagcagcagcagcagcagcagcagcagcag 1239
Oy 1342 aagactaatggaagagaatctgcaagcgtatctcgaagagatgtagatgtagatgtagatg 1401
Db 1240 cggaaacgcaaggtgtagcagagtgtagatgtagatgtagatgtagatgtagatgtagatg 1299
Oy 1402 ttgattgcaaatctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1460
Db 1300 ctggaagagtgcccgctccacgagtagatgtagatgtagatgtagatgtagatgtagatg 1358

```

RESULT 15

AAAF74190

ID AAF74190 standard; DNA; 1280 BP.

AAAF74190:

02-MAY-2001 (first entry)

DNA encoding environmental stress tolerant protein SEQ ID 7.

```

KW Environmental stress resistance; salt; heat; desert; transgenic plant;
KM ds.
OS Bruguiiera sezanigula.
PN WO200106006-A1.
XX 25-JAN-2001.
XX 19-JUL-2000; 2000WO-JP04862.
XX 19-JUL-1999; 99JP-0235910.
XX 24-MAR-2000; 2000JP-0085377.
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA Yamada A, Ozeki Y, Saito T;
PI
XX
XX
XX MPI; 2001-147355/15.
DR P-PSDB; AAB80611.
PT Screening method to obtain DNA encoding environmental stress resistance
PT factor, useful for producing transgenic plants resistant to
PT environmental stress -
PS Claim 23; Page 84-86; 167pp; Japanese.
XX
XX Polynucleotide sequences AAF74187 - AAF74218 encode proteins
CC AAB80608 - AAB80639, which impart environmental stress resistance. The
CC invention relates to a method for identifying DNA encoding proteins
CC imparting environmental stress resistance. The method comprises inserting
CC cDNA from a library originating in a salt-resistant organism into a host
CC cell, culturing the transformants under conditions in which the
CC untransformed host does not grow well, and selecting for viable clones.
CC The method is useful for obtaining DNA encoding environmental stress
CC resistance factors. The DNA encoding proteins conferring environmental
CC stress resistance, can be used in the production of plants resistant to
CC environmental stress, which can be cultivated in unfavourable
CC environments such as deserts, salt damaged ground, cold regions and the
CC oceans. They can be used for increasing the area of land covered by green
CC plants, and desert greening and afforestation, in order to counter the
CC effects of the increase in atmospheric carbon dioxide concentration. PCR
CC primers AAF74219 and AAF74220 are used in an example illustrating the
CC method of the invention.
XX
XX Sequence 1280 BP; 388 A; 231 C; 344 G; 317 T; 0 other;

```

Query Match 8.7%; Score 166; DB 22; Length 1280;

Best Local Similarity 51.9%; Pred. No. 1,1e-28; Matches 444; Conservative 0; Mismatches 385; Indels 27; Gaps 2;

```

Oy 605 agttagatttcaattatgattgattgattgattgattgattgattgattgattgattgattg 664
Db 110 atttgagggggaaatagatagagagcattggggaagaaagtgcaagtgaggatgtagta 169
Oy 665 tatacatcgaaagcaaatagtgagcagtgagaaagagtgtagatgtagatgtagatgtagatg 724
Db 170 ttgcaattgcaagagtgtagcagcaaatatcaaaagcttgtagcattatcttcgtagtca 229
Oy 725 cagaatagat-----ctggaagctggaagagtagtatttctcaccacaagtgtagtcc 778
Db 230 gggattacgattgcatgtaggacacagtgtagatgtagatgtagatgtagatgtagatgtagatg 289
Oy 779 atagaanaaagaatgtgcaagatgtcacacttaccattgtagtgaagaacatgtagtcc 838
Db 290 agaaagagaaagagtgtagatgtagatgtagatgtagatgtagatgtagatgtagatgtagatg 349
Oy 839 agccacaagtgtagcagaatattgtccctatgtagcagatgtagatgtagatgtagatgtagatg 898
Db 350 gaacacaggggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 388
Oy 899 ctgaatcaccgaaataactacgcaagaataataagtgtagatgtagatgtagatgtagatgtagatg 958

```

Db	389	gfgaaatccfvgcgagggtgagggaaacaaattgacccaagaagtgtgctgaatvgagaaagg	448
QY	955	aaggaattgcagagcttgctgaacttggtgttttggatctgatgagtcaccaatgtgtga	1018
Db	449	aagggaagaacagagaaattgtgtccaaagtgtccctcttttatgatgagtgccacatggttgaca	508
QY	1019	tgcgaattgttttcttactacttaacccgtgcattgagagcccatattaccacatcgtgaac	1078
Db	509	ttagtgcttcttcaattctctgaatccggtcctcttgagaaatgagatggtcgccaataattagttg	568
QY	1079	tgtgcacaataagvgggaataatgtaatgtaaagagaactgatatagtcaaatgccacatgta	1138
Db	569	tgtgcacaacaagvgggaatcaaccaacaatcagacgaaccaaatattacaactctcctaattgga	628
QY	1139	taccggtggaatcctcttagatagctgtgtgaattattctcgacagagaacatattgcccactgt	1198
Db	629	tccaataagatcctcttgatctgacactactcaattacccaactlaaactlaccaagaagtgt	688
QY	1199	agatgataacacataatltgctatcccgacccaagtgaagagatgtatatgtatgaaga	1258
Db	689	aaatcgtgaagatctctgtgatatcagaatgtgcagaagaagaatgtgtgagaatgctctaaagg	748
QY	1259	gtcttgcttatttaggcgagatcgcgaacgcgaacgaacatctttgagacatgctatccaatgta	1318
Db	749	caaaagctgtgtttaaacaacataattgggacagaaacatcccttgagataatgcatccatccaca	808
QY	1319	tatcacctgcgcgaaggtgtgcttcaaaagacaaatgaagagagaagaatctgcgaagctgac	1378
Db	809	ttaactgttcagcatctggcatgcatgccaaagcgaagaaggaagactgtgtgaaactgaagaca	868
QY	1379	tccaggaagtcagtggtgctctatatttgatggccaacatccctcggtcgtgcttcccaagg	1438
Db	869	ttagtcgagcttacaatctgtttcttgatgtaaagaagatctacacagtaacttaataagt	928
QY	1439	aacagaagaagatacat	1454
Db	929	atcagaatcagtaact	944

Search completed: November 13, 2001, 11:55:18
Job time: 4368 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2001, 09:53:20 ; Search time 2536.09 Seconds
(without alignments)
11661.391 Million cell updates/sec

Title: US-09-589-510-3

Perfect score: 1912
Sequence: 1 acccagcgctccgaattt.....aaaaaaaaaaaaaaaaaaaaa 1912

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_p11:*
13: gb_p12:*
14: gb_p13:*
15: gb_p14:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_p1:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_hc91:*
61: gb_hc92:*
62: gb_hc93:*
63: gb_hc94:*
64: gb_hc95:*
65: gb_hc96:*
66: gb_hc97:*
67: gb_hc98:*
68: gb_hc99:*
69: gb_hc10:*
70: gb_hc11:*
71: gb_hc12:*
72: gb_hc13:*
73: gb_hc14:*
74: gb_hc15:*
75: gb_hc16:*
76: gb_hc17:*
77: gb_hc18:*
78: gb_hc19:*
79: gb_hc20:*
80: gb_hc21:*
81: gb_hc22:*
82: gb_hc23:*
83: gb_hc24:*
84: gb_hc25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	1910.8	99.9	1912	10	AX077226	AX077226 Sequence
2	1686.6	88.2	1845	10	AX077224	AX077224 Sequence
3	1630	85.3	1869	10	AX077232	AX077232 Sequence
4	1362.2	71.2	1886	10	AX077228	AX077228 Sequence
5	1351.8	70.7	1898	10	AX077230	AX077230 Sequence
6	827.8	43.3	1755	14	CAR276264	AJ276264 Cicer a
7	631.4	33.0	1638	94	BC004718	BC004718 Mus mus
8	631.4	33.0	1709	94	AF100694	AF100694 Mus mus

9	622.6	32.6	1567	94	AB001581	AB001581 Rattus norvegicus
10	622.6	32.6	1587	10	E21815	E21815 Protein for
11	622.6	32.6	1587	94	AB002406	AB002406 Rattus norvegicus
12	615.6	32.2	1371	8	AF218072	AF218072 Xenopus laevis
13	612.4	32.0	1710	85	AB012122	AB012122 Homo sapiens
14	612.4	32.0	1730	10	E21816	E21816 Protein for
15	612.4	32.0	1733	92	HSA010058	HSA010058 Homo sapiens
16	612.4	32.0	1736	88	AF099084	AF099084 Homo sapiens
17	612.4	32.0	1746	91	BC002993	BC002993 Homo sapiens
18	612.4	32.0	1750	88	AF070735	AF070735 Homo sapiens
19	612.4	32.0	1789	91	HOSA18419	HOSA18419 Homo sapiens
20	612.4	32.0	2503	91	BC000051	BC000051 Homo sapiens
21	572.6	29.9	1681	5	AF233278	AF233278 Drosophila
22	534.6	27.9	1510	9	AX011662	AX011662 Sequence
23	483.6	25.3	19236	14	SC9346	Z48784 S.cerevisiae
24	444.6	22.3	11108	63	AC015338	AC015338 Drosophila
25	444.6	22.3	138615	60	AC008139	AC008139 Drosophila
26	444.6	23.3	221888	5	AE003686	AE003686 Drosophila
27	417.4	21.8	953	53	CNS06HCM	AL398828 T7 end of
28	408.4	21.4	110000	84	LMF1CHR4_0	AL459663 Leishmania
29	380	19.9	915	53	CNS06PPS	AL040904 T7 end of
30	354.2	18.5	24775	5	CCE27H6	Z81042 Caenorhabditis
31	346.6	18.1	41648	5	GERG38K16	AC084520 Caenorhabditis
32	341.6	17.9	964	53	CNS06Z38	AL471838 T7 end of
33	326.2	17.1	1105	53	CNS06LX8	AL404754 T3 end of
34	305.8	16.0	250203	6	CNS07EG1	AL590451 chromosome
35	296.2	15.5	860	72	AC052282	AC052282 Giardia
36	276.6	14.5	940	71	AC039246	AC039246 Giardia
37	276	14.4	1389	8	AF218071	AF218071 Xenopus laevis
38	268.2	14.0	863	53	CNS06Z02	AL421704 T3 end of
39	263.2	13.8	37235	15	SPBC63	AL035536 S.pombe
40	263.2	13.8	110000	84	PFMA13P2_0	AL049185 Plasmodium
41	256.2	13.4	293350	2	CNS06PX02	AL242884 Pyrococcus
42	256.2	13.4	349980	9	AX041918	AX041918 Sequence
43	252	13.2	128310	77	AC087325	AC087325 Trypanosoma
44	246.2	12.9	643	53	CNS07917	AL454545 T3 end of
45	245.8	12.9	987	53	CNS06K0N	AL403149 T7 end of

ALIGNMENTS

[illegible]

RESULT	1				
LOCUS	AX077226	1912 bp	DNA	PAT	22-FEB-2001
DEFINITION	Sequence 3 from Patent WO0105975.				
ACCESSION	AX077226				
VERSION	AX077226.1	GI:13121820			
KEYWORDS	.				
SOURCE	Zea mays.				
ORGANISM	Zea mays. Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta Magnoliopsida; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 1912) Mahajan, P. B. Maize orthologues of bacterial ruvB:cdnas and uses thereof Patent: WO 0105975-A 3 25-JAN-2001; PIONEER HI-BRED INTERNATIONAL, INC. (US) Location/Qualifiers 1..1912 /organism="Zea mays" /db_xref="taxon:4577" 94..1461 /note="unnamed protein product" /codon_start=1 /protein_id="CAC32389.1" /db_xref="GI:13121821" translation="MRTVEVSTKSKORIATHTTHIKGIGDANGMALALAGVCGOAA AREAGGLAVDMIRQKMGRAVLGAPGATGKTALAGIAGIOLGSKVPCFPMVSGEWA LSEVKEVPEVNRFRALGRIKENKEVGEVTELSPEAESTGGVAKSISHVILI LKYAKGQKLDSSTYDALIKEVYANGVYIIPANGAKVARGRCDSRTEYEDLEA EVPYIPKCEVHKKEIVQDVTLHOLDANAPOGGDIISLMCGQMKPKRTETTELK				

QY	1366	tcgaagctgatactcaggaagtcagttgggctctattggatgccaatcctcgctcg	1425
Db	1285	tcgaaggctgatactcaggaagtcagttgggctctattggatgccaatcctcgctcg	1344
QY	1426	ctgcctccaggaacaaagaagaatacaatacctaagtttgatctccctgctggtaagt	1485
Db	1345	ctgctccaggaacaaagaagaatacaatacctaagtttgatctccctgctggtaagt	1404
QY	1486	ctcgaagaagatgtagtttcacagctcgaagtcacatctagtcgtatctgctccacag	1545
Db	1405	ctcgaagaagatgtagtttcacagctcgaagtcacatctagtcgtatctgctccacag	1462
QY	1546	ttcatagctactgctgctcttggagcgagacaaattcgagggggaacggctgaaatttgag	1605
Db	1463	-----gctcttgagcgagacaaatttcggggggagcgcttgaaatttttcag	1509
QY	1606	tgcctgctgtgtagtactccaragaagacttggtccgcgaatgctcgttcaagac	1665
Db	1510	tgcctgctgtgtagtactccaragaagacttggtccgcgaatgctcgttcaagac	1569
QY	1666	tgttcgctgattagattggtcacccggtgacgaagtgcgcgtgtgtg-itttatactgct	1724
Db	1570	tgttcgctgattagattggtcacccggtgacgaagtgcgcgtgtgtg-itttatactgct	1629
QY	1725	catcggtctccggagct--tgcctccacgggtgtgtatgtggccggaacctatcttggaa	1782
Db	1630	catcggtctccggagcttgcctccacgggtgtgtatgtggccggaacctatcttggaa	1689
QY	1783	ccatgataatagatagattctctacagaatgcaactgcatgcttattatcttaaa	1842
Db	1690	ccatgataatagatagattctctacagaatgcaactgcatgcttattatctttaa	1749
QY	1843	tgctcataagacataacgaaatgcttctacaacmcttaaa	1882
Db	1750	tgctcataagacataacgaaatgcttctacaacmcttaca	1789
RESULT	4		
AX077228	1886 bp	DNA	PAT
LOCUS	AX077228	1886 bp	DNA
DEFINITION	Sequence 5 from Patent WO0105975.		22-FEB-2001
ACCESSION	AX077228		
VERSION	AX077228.1	GI:13121822	
KEYWORDS	.		
SOURCE	Zea mays.		
ORGANISM	Zea mays.		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta		
	Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;		
	Panicoideae; Andropogoneae; Zea.		
	1 (bases 1 to 1886)		
	Mahajan, P. B.		
REFERENCE	Maize orthologues of bacterial rnbv:cdnas and uses thereof		
AUTHORS	Patent: WO 0105975-A 5 25-JAN-2001;		
TITLE	PIONEER HI-BRED INTERNATIONAL, INC. (US)		
JOURNAL	Location/Qualifiers		
FEATURES	1..1886		
source	/organism="Zea mays"		
	/db_xref="taxon:4577"		
	82..1449		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC32390.1"		
	/db_xref="GI:13121823"		
	/translation="MRTIEVQSTSKRQIATHTHIGLGLDANGSMPLAAGFYGOAA		
	AREAGLAVDMIRQKMMAGRALLAGPATGATLALGTAEIGSKVPCPMWGEVAV		
	SSEYKTEVLENERRAIGLRKENKEVGVGEVETSPREASTGTAKSISHTYIG		
	LKTVAGKQLKDSITVDALIRKEVAVGVITYTEANSKAVKYGCDSTAFETDEAE		
	EYVPIRKEVYKKKEIYQDVTLDHLDHANAQPGGDDILSLGQMMKPRKTEIEKLE		
	QELKNVNRITDESIALVPGVLFIDEVHMLDECHSYLRALESPISPIVLIATNNG		
	ICNNKQDNTSPHGIIVDLIDRLVYIIRTEYGTETMIQTLAIRQVEIDIDESTLAW		
	LGELIGQDTSLRHAIQILSPASVYAKTNGREKKMKADLEVSGLYLDAKSARLLQEQQ		
	ERYLT"		
CDS			

QY	728	tgaagctgaagagatgatgtlccctatcccacaagatggaagtcgaataaagaagaattgt	797
Db	813	TGAAGCTGAAGAGATGATGTTCTATCCCAAAAGGTGAAGTCATTAAGAAAGAAATAGT	872
QY	798	gcagaatgcaacactcattactcttgaatgcagcaaatctgaagccaaaggtgtgcaga	857
Db	873	GCAGAGTGCACACTCCATGACCTTGTATGACCAAAATGCCCCAGCCACAAAGGTGGCCAGA	932
QY	858	tatttgtccctctatgggccagatgtgaacccacgaagaactgaactcaccgaanaact	917
Db	933	TATTTGTCCCTTATGGGCCAGATGATGAAGCCACGGAAGACTGAAATCAACCGAAAGCT	992
QY	918	accgcaagaataatgaaggtggtaatatagatataatcgaatgaaggaatgcaagcttgt	977
Db	993	ACCGCAAGAATAATGAAGGTGGTAAACAGATATATCCACMAAGAAATCCACAGCTTGT	1053
QY	978	accbgtgttttgttcatatgatgaagtcacaatgtgtgatatacgaatggtttcttctatc	1033
Db	1053	ACCTGGTGTTTTGTTCATTGATGAGGTCCACATGTTGGATTTTAAATGCTTTTCTTATCT	1113
QY	1038	taacgttgtatttggagagcccatctacacaacatcgtgatactgtgtctacaatatggggaat	1093
Db	1113	TAAACGGCATTTGGAGAGCCCATTTATCCAAATGTGATACCTGCTACGAATAGAGGAAT	1173
QY	1098	atgtaatgtaaagaggaacatgatatacaagtcacaatgatataccgttgtatctctaga	1153
Db	1173	ATGTATGTGAGAGGAACGATATGACGAGTCCACATGTGATACCATGATGACACTTCTTACA	1233
QY	1158	taagctgtgatattctggacagagacatatgtgccactgagatgatalacagatatgtgc	1213
Db	1233	TAGGTTGGATATTATTCGGACAGAAACATATAGGCCCTTACGTAGATATACATGACTGTGCG	1293
QY	1218	taccgaagacacaagtgtgaagagatgtgatgtatgtatgaaagaagcttgatattatgaaga	1277
Db	1293	TATCCGACGACAAGTGGAAAGATTTGATATGATGATGAAGAAAGCTTGCTTATTTAAGGCGA	1353
QY	1278	gattcgcgaacagacacattcttggagacatgcatatcaatltgatatacactgtccagactgtgt	1337
Db	1353	GATCGGACAGCGACATCTTTGAGACATGCTATTCATGTTGCTATTCACCTGCCAGCGTGT	1412
QY	1338	ctcaagaagctaatgtgaagaagaanaactctgaagcctgcatctctggagaagtaagtgtgct	1397
Db	1413	CCCAAAAGCACAGGGAGAGAAAAGATGTGCAAAGCTGACCTCGAAGAGTACAGCGGGCT	1472
QY	1398	ctatttgaatgccaatctcctgcgtcgtgcctgtctccagagagacaaagaagaatatacatc	1457
Db	1473	CTATTTTGATGCAAAATCCTCGCTCGCTCGCTGCCAGGAGCAACGAAGAATATCATCAC	1532
QY	1458	ctagatttggatctccctcgtctgtgtgaagctccgaagaagatgtagtgtgcagctgaaagt	1517
Db	1533	CTAGACTTGCAATCTTCCTGCTGTGGAAAGAAAGCTCGAAGAGAAATGATCTTCCAA--	1590
QY	1518	catctaatgtcatatcatctgtcttccacaggttcatcactaactgtgtcttgaagcgaacat	1577
Db	1591	-----AGTGCCTCATCTGCTTCAGTATGCTGATGCTCAAGTCTTGGAAGAGACACGT	1645
QY	1578	tttcggggggg-----aagcgctgaatttttgaagtgctgt-ctgtgttatgtctccara	1629
Db	1646	TTTCTCGGGGGAAGAAACAGTGTGAAATTTTGTAGTACCTGTTTTTTGTATGACCTCCAGC	1705
QY	1630	gaagactgtgttccgacataatgtcgtctcaacgaactgtctgcctgattaga-----tgtgt	1685
Db	1706	GAAAGCTTGATTTCCGCATATATGCTTGTTCAGCACCGTTAACACCGTTAAGATTGGCC	1765
QY	1686	caacggtgacaggaatgtgcgtgtgtgttttla-----lctgtctcaatcgtgtgtccgaact	1742
Db	1766	CACCGGTGCAGGAATTTCCGGGTGTGTTTTTCAACCCCTCGCTAAATCGATGTCGCAATCT	1825
QY	1743	gctcccaacggttgtatgtgcgcccaaacactatcttgtacaatcattgata 1791	
Db	1826	GTGTGCCACAG--GGTGTGTATGTAGACCTACTCTTTGTATACCATGCACA 1872	


```

QY 1171 attcgagacacatatgccctactgagatgatacagatatattggctatccgagacaa 1230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1152 ATCAGGACCATGTTATACAGCCACAGGAGATGAGCAGATATTAAAGTCCGAGCCAG 1211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1231 gtgagagattgatgatgaagaaagtctgttatttatttggcagatcgacagacag 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1212 ACGGAAGGCATCAACATCAGCAGGAGGCGCTGAACCACTCGGCGAGATTGGCACCAG 1271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1291 acatctttgagacatgctattcaattgatatcacctgccagcggtgtctcaagactaat 1350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1272 ACCACTCTGAGGTATTCGGTGCAGCTGCTGACCCAGCCAGCAACCTGCTGCAAGATCAAT 1331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1351 ggaagagagaaatctgcaagctgctatcgcaggaagtcagtggtctatttggatgcc 1410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1332 GGCAGGACAGCATTTGAGAGGAGCAGCTGGAAGAGATCAGGAGCTCTCTATGACGCC 1391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1411 aaatcctcgtcgtcgtctccaggagcaacaaagaaagatacacatccttagattgg 1467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1392 AAGTCTCTCCCAAGATTTTGGCTGACAGCAGGACAGTACATGAAGTAAGCTTGG 1448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AF100694 1709 bp mRNA ROD 05-JAN-1999
LOCUS Mus musculus Pontin52 mRNA, complete cds.
DEFINITION AF100694
ACCESSION AF100694
VERSION AF100694.1 GI:4106527
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1709)
AUTHORS Bauer, A., Huber, O. and Kemler, R.
TITLE Pontin52 is ubiquitously expressed during mouse embryogenesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1709)
AUTHORS Bauer, A., Huber, O. and Kemler, R.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1998) Department of Molecular Embryology,
Max-Planck Institute of Immunobiology, Stuebeweg 51, Freiburg
D-79108, Germany
FEATURES
source
1. 1709
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="C57MG"
/tissue_type="mammary glands"
/note="expressed during mouse embryogenesis"
145..1515
/codon_start=1
/product="Pontin52"
/protein_id="AAB02877.1"
/db_xref="GI:4106528"
/translation="MKIEEVKSTTKTORIASHSVKGLGDSGLAKOAAAGLAVGOEN
ARACGVIVELISKMGRAVLVLPAGPGTKTALALATAOELGSKVPFCPMWGEVY
STEATKTEVLMENFRRAIGLRIKETREVEGEVTELTPTCEINPMGGYKGTISHVIG
LTKATKPKQLDPSIPESQKERVEAGDVIIYIANSVAVKRGKRDYATVETDLEAE
EYVPLPKGDVHKKEIITQDVLHDLVDVANRPOGQDILSMGQMLMKPKKTEITDKLR
GEINKVNKYIDOGVABLVPGLFVDEVHMLDIECTYILHRALESSIAPIVIFASNRG
NCVTRGTEITLSPGIPDLIDLRVMIIRTYLTPQEMKOIKITRAQTEGINISEALN
HLGEIGTKTYLRVSVQLLTIPANLLAKINGKDSIEKEHEVEISELSELFYDKSSAKILADQ
ODKYMK"
BASE COUNT 480 a 388 c 502 g 339 t
ORIGIN
Query Match 33.0%; Score 631.4; DB 94; Length 1709;
Best Local Similarity 66.7%; Pred. No. 8.3e-108;
Matches 918; Conservative 0; Mismatches 456; Indels 3; Gaps 1;
QY 94 atgagatcgagaggtgctcagctcagagagcagcgcatcgccacccacacacac 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 145 ATGAAGATTGAGGAGGTGAAGAGACACCAAGAGCAGCGCATCGCCTCCACAGCCAC 204
QY 154 atcaaggagactcgcctcgcagcccaatgggatggcgattcgttggcggcggggttgcgtg 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 GTGAAGGGTCTGGGCTGGAGCAGAGCGGCTAGCCAAGCAGGCGGCTCGGGTCTCGTG 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 214 ggcagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 GCCCAGGAGAAACGCGCAGAGAGGCATGTGGCGTCAATTAGAAATTAATCAAAACCAAGAAA 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 274 atgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 ATGGCTGGAAGAGCTGCTTCTTGGCGGGGCTCTCTGGAACCTGGCAAGACAGCCTTGGCC 384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 334 ctcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 CTGTCTATTGCTCAGGAGCTGGCAGTAAAGTCCCTTCTGCCCGATGGTGGTAGCGAG 444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 394 gtgtactcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 445 GTATACTCACTGAGATCAAAAAGACAGAGGTGTGATGGAGAACTTCCGAGGGGCCATT 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 454 ggtttgctataaaaggaacaaagaggtttatgaagagaggttactgaaactttcccca 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 505 GGGTTGCCGATTAAGGAGACTAAGGAGGTTTATGAGGGGAGGTGACAGAGCTCCTCCTCC 564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 514 gaagagcgtgaggtacaaactggttgatgatcgaagagcattaccatgtaatacatcagc 573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 565 TGTGAGACAGAGAACCCCATGGTGGTACGCCAAATATCAGCCATGTGATCATAGGG 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 574 ttaagactgttaaaggactaaagcaactgaagttagattctcaatttatgctctg 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 625 CTCAAGACTGTCTAAAGGAACCAACAGCTGAAGCTGGACCCAGTATTTTGAAGTTG 684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 634 atcaagaaaaggtgagcgtggtgatttatatacatcgaagcaaatagtgaggcagtg 693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 685 CAGAAAGAACGAGTAGAGGCTGGAGATGCTGATTTACATTGAAGCAAAATAGTGGAGCTGTG 744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 694 aaaagagtgtagatgtagtcttcttctcagagaatcagatcctgaagcgaagagat 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 745 AAGAGGCAAGCAGGTGTGACACCTATGCCACAGAGTTTGACCTTGAACCTTGAACCTTGAACAGAGTAC 804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 754 gttctatcccccaaggtgaggtccataagaagaaaattgtgcagagatgcacact 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 805 GTCCCTTTGCCAAAGGAGAGATGTGCACAGAAGAAATCATACAGGATGTGACCTTG 864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 814 catgacctgtatgcagcaaatgtcagcccaaggtgcccgaagatatatttgccttatg 873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 865 CATGACCTGGATGTGGCTAATCGCGGCTCAGGGTGGGCAAGATATTTCTGTCTATGATG 924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 874 gcccagatgataaacaccagaagactgaatccccgaaactacgcgaagaaataat 933
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 925 GGCCAGTGTGAAGCCAAAAGACAGAGATCAGAGTAACTTCGAGGGGAGATCAAC 984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 934 aagggtgtaataatagatatatcgatgaagaaattgcagagcgttgacctggtttgttc 993
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 985 AAAGTGTGAACAAATACATTACCCAGGCTGTGCGCGAGCTGGTCCCTCGAGTGTCTTTT 1044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 994 attgatgagtcacatgttgatgatcgaatgttttttcttataccttaacgcgtgag 1053
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1045 GTCGATGAGGTCCACATGCTGATATCGAGTCTTTACCTACTCTCAGCCAGCCCTGGAG 1104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1054 agcccaattatcaccaatcgtgacttctgatacaaataggggaatatgtaataagagga 1113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1105 TCCTCCATCGCCCCCATTTCTCATCTTTCATCCAACCGAGGCAACTGTCTCATCAGGGGC 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1114 act---gatagacaagtccacatggtatataccggttgatcttcttagatagcgtgtgatt 1170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1165 ACTGAGGACATCACTTCTCCACAGGCATCCCGTTGGATCTCTGTAGACAGGTGATGATC 1224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1171 attcgacagagacatatggcctactgagatgatacagatatattggctatccgagacaa 1230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1225 ATCAGGACCATGCTATACAGGCCACAGGAGATGAAGCAGATCATTAAGATCCGAGCCAG 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

BASE COUNT	432 a	360 c	457 g	318 t
ORIGIN				
Query Match		32.6%	Score 622.6;	DB 94; Length 1567;

QY	94	agggagatcagggaggtgcaagtcgcagcttcgaagaagcagcagtcgcacccacacccac	15
Db	18	ATGAAGATTGAGGGAGTGAAGAGACACCAACCAACACCAACGCAATTGGCTCCACAGCCAC	77
QY	154	atcaaggagtcgcgccttcgcagcccaatggatggatggatggatggatggatggatggat	21
Db	78	GTGAAGGGGGCTGGGGCTGGATGAGACAGCGCTTGCCACAGCAGCGGCTTGGGGCTCGTG	13
QY	214	ggccaagc	27
Db	138	GGCCAGGAGAAACCGAGAGAGGCGCATGTGGTGTCAATGTGCAATTATCAAAAGCAAGAA	19
QY	274	atggccgc	33
Db	198	ATGGCTGGAAAGACTGTCTTTGTTGGCAGGGGCTCTCGGAACGTGGCAACACACCCCTTGGCC	25
QY	334	ctgcgcatagcaccaagagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	39
Db	258	CTGGCTATTCTCTCAGGAACCTGGGCACATAAGTCCCTTTTGCCCATGTGGGTATGTCAA	31
QY	394	gtgtactcctcgc	45
Db	318	GTTATCTCAACTGAGATTCAGAAAGACAGAGGCTGTGATGAGAACTTCCGAAGGGCTATT	37
QY	454	ggttcgcgtataaaggaagaaacaaagagtttttgaaggaagaggttactgcgaacttccca	51
Db	378	GGGCTCGCGATTAAAGAGACTTAAGAGGTTTATGGAAGGGAGGATGACAGACTCACTCC	43
QY	514	gaagagcgtgaagatlaacaactcgttgatatacgcaaaagcatltagcattgatacga	57
Db	438	TCTGAGACAGACAGACCCCATGTGGTGGTATGGCAAACTATCAGCACTGATCATAGG	49
QY	574	ttaaacactgtaaaggagactaagcaactgaagttagatttccaattatgatgctcgt	63
Db	498	CTCAAGACTGCCAAAGAACCAAAAGCTGGAAGCTGACCCAGCTATTTTTGAAGAGTTTG	55
QY	634	atcaagaaagagtgccagtggttgatgttatacatcgaagaaatagtlgaagcagtg	69
Db	558	CAGAAAGACGAGTAGAGCTGGAGATGTGATTTCATTGAAGCAACAGTGGAGCGCTGTG	61
QY	694	aaaagagttgtagatgtgattccttctgctacgaatacagctctgaagctgaagat	75
Db	618	AAGAGGCAAGGCGAGTGTGACACCTATGCTCCACAGAGTTTACCTTGAAGCTGAAGATAT	67
QY	754	gttccatccccaaagggtgaagtcataatgaagaaagaattgtgcagatgtcacatt	81
Db	678	GTCCCTTTGGCCAAAGGAGATGTGCACAGAAAGAAATCAATACAGATGTGACCTTG	73
QY	814	catgacattgatgcagcaaatcgtcagcccaagaagtgccgaagatattgttccattg	87
Db	738	CACGACTTGGAGGTGGCCAAATGCCGGGCTCAGGGTGGCGCAAGATATTCTGTATTAGATG	79
QY	874	ggccaagatgatgaacccaagaaagacttgaataccgcgaaaaactacgcgaagaatlaa	93
Db	798	GGCCAGTTGATGAAGGCCAAAAGAACAGACAGATCCACATTAACCTTGAGGGGAGATCAAC	85
QY	934	aaggtgtgaatatagatatatcgaatgaaggaattgcagaagcttgtaccgtgtgtgtc	99
Db	858	AAGGTGGTGAACCAATATCAATTGACCAAGGGGCTTGCACAGCTGGCTCCCTGAGAGCTTT	91
QY	994	attgatgggtgcacatgttggatagatgaattgtttctatcttbaagcgtgcattggag	105
Db	918	GTTGACGAGGTCCACATGTGAGATGTGATGATGCTTTACCTACTGCACCGAGCCCTGGAG	97
QY	1054	agcccatattacaacatcgtgatacttgtacaatataggagaatagtatgtlaagagga	111
Db	978	TCTTCACATGGCCCCCATTTGTCAATTTTGTGATCCCAACCGAGGCAACTGTGTCAATGAGGGC	103
QY	1114	ac---tatatgacaagtcacacatgatatccgggtgagatctctagatagagctgtgatt	117

[illegible]


```

QY 1114 act---gatatgacaagtcacatggtatataccggtgagatcttcttagataggtgtgatt 1170
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1080 ACTGAGGACATCATCCCTCAGCGCATCCCTCTTGACCTTCTGGACCGAGTGATGATA 1139

QY 1171 attggacagagacatagccctactgagatgatatcacatattggtctatccgagacaa 1230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1140 ATCGGACCATGTGTATCTACAGGAAATGAAACACATCATTAATAATCCGTCGCCAG 1199

QY 1231 gtgagagagatgatatgagaaagcttctgctattttagcgagatcgagacag 1290
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1200 ACCGAGGAATCAACATCAGTAGGAGGCACTGAACCACTGGGGAGATTGGCAACCAAG 1259

QY 1291 acatctttgagacatgctattcaattgatcatcctgcagcgctggtctcaagactaat 1350
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1260 ACCACACTGAGGTACTCAGTCAGCTGTGACCCCGGCACTTCTTGTGCTAAATCAAC 1319

QY 1351 ggaagagagaaatctgcaagctgacatcgcgaggaagtcagtggtctctatttgatgcc 1410
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1320 GGAAGAGCAGCAATGAGAAAGAGCATGTGCAAGAGATCAGTGAACCTTTCTATGATGCC 1379

QY 1411 aaatcctcgctcggtctgctccagagagcaacaagaatacatca 1456
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1380 AAGTCTCGCCAAATCTGTGCTGACCAAGCAGGATAGTACATCA 1425

RESULT 15
HSA010058 HSA010058 1733 bp mRNA PRI 20-NOV-1998
LOCUS Homo sapiens mRNA for NMP238 protein.
DEFINITION AJ010058
ACCESSION AJ010058
VERSION NMP238 protein.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1733)
Holzmann,K.
Direct Submission
Submitted (03-AUG-1998) Holzmann K., University of Vienna,
Institute of Tumor Biology-Cancer Research, Borschkegasse 8a,
Vienna, A-1090, AUSTRIA
2 (bases 1 to 1733)
Holzmann,K., Gerner,C., Korosec,T., Poltl,A., Grimm,R. and
Sauermaun,G.
Identification and characterization of the ubiquitously occurring
nuclear matrix protein NMP 238
Biochem. Biophys. Res. Commun. 252 (1), 39-45 (1998)
990332802
FEATURES
source Location/Qualifiers
1..1733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE 591284"
/tissue_type="pancreas"
77..1447
/function="Related to AAA family, role in ATP-dependent
anchoring"
/codon_start=1
/product="Nuclear matrix protein NMP238"
/protein_id="CAA08986.1"
/db_xref="GI:3892584"
/translaton="MKIEVKSTTKTORIASHHVHKGLDESLAKQAQASGLVQGEN
AREAGVIVELIKSKMAGRAVLGPPGKTALALIAOELGSKVPFPMVGSSEVY
STEIKKTVLMENFRAIGLIKETKEVEYELTPCETENPMGSGYKTIISHVIIG
LKTAGTKOLKIDFSIFSLQKERVEADVIYEANSKAVKROGKIDYATEFDLEA
EYVLPKGDVHKKEIIQDVTLDHLDVANRPOGGDILSMGOLMKPKTETDKLR
GEINKVKNYIDOGIELVGVLFVDEVMLDIECTPYLHRALESSTAPIVIFASNRG
NCVIRGTEDITSPHGIPLDLDLDRVMIIRTMLYTPQENKQIKIKRAQTEGINTSEALN
HLGELGKTTILRYSVQLITPANLIANKDSTIEKHEVEISELFDYDAKSSAKILAQ
ODKYMK"
1733
polya_site

```

```

BASE COUNT 462 a 403 c 481 g 387 t
ORIGIN

```

```

Query Match 32.0%; Score 612.4; DB 92; Length 1733;
Best Local Similarity 66.0%; Pred. No. 2.8e-104;
Matches 902; Conservative 0; Mismatches 461; Indels 3; Gaps 1;

QY 94 atgagatcgagaggtgagtcagtcacctggaagaagcagcgcacccacacacacac 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 ATGAAGATTGAGAGGTGAAGAGCAGCTAGAAAGACGAGCGCATCGCTCCACAGCCAC 136

QY 154 atcaaggactcgccctcgaccccaatggatggcgattggttgcgcggggttcgtg 213
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 GTAAAGGGCTGGGCTGACAGAGCGGCTTGGCCAAAGCGGCGCTCAGGGCTTGTG 196

QY 214 gcccagcgccgcccgcgagggcgccgggtggcgctgacatgattcgccagaagaag 273
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 GCGCAGGAGAACCGCGAGAGGCATGTGGCGTCATAGTAGAATTAATCAAAAGCAAGAA 256

QY 274 atggccggccgcccgcggtgctcttcggttcgcccgcgagcagcgagcgagcg 333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 ATGGCTGGAAGAGCTGTCTTGTGGCAGGACCTCTCTGGAAGACTTCCGACAGGCATT 316

QY 334 ctccgcatgccagagctcgccagcaggtcccttctctctctatggttagatcagaa 393
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 CTGGCTATTGCTCAGAGCTGGTAGTAAGTCCCTCTCTGCCCAATGGTGGGAGTGAA 376

QY 394 gtgtactcctcgaggttcaagaaaactgaggtgctgtgtagaaatttccgttagagctata 453
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 GTTTACTCACTGAGATCAAGAAGACAGAGGTGCTGTGATGAGAACTTCCGACAGGCATT 436

QY 454 ggttgcgtataaaggaaacaaagaggtttatgaagagagattactaaacttccca 513
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 437 GGGCTGCGCAATAAAGGAGACCAAGGAAGTTATGAAGTGAAGTCACAGAGCTTAACTCG 496

QY 514 gaagagcgtgagatcaaacactggtggatgataaagaagcattagccatgtaatacacc 573
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 497 TGTGACAGACAGAAATCCCTGGGAGGATATGGCAAAACCATTAGCCATGTGATCATAGA 556

QY 574 ttaagactgttaaaggagactaagcaactgaaagttagattctcaatttatgatgctctg 633
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 557 CTCAAACAGCCAAAGGAACCAACAGTTGAAACTGGACCCCGACAGCATTTTGAAGATTG 616

QY 634 atcaaggaaaagggtggcagtggtgattgtatatacatcgaaagcaaatatgagcagctg 693
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 CAGAAAGAGCGAGTAGAAGCTGGAGATGTGATTTACATTGAAGCCCAACAGTGGGCGCGTG 676

QY 694 aaaagagttgtagatgtagtattctttgtacagaatcagatcttgaagctgaagagtat 753
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 AAGAGCGAGGCGAGGTGTATACCTATGCCACAGAAATTCGAGCTTGAAGCTGAAGAGTAT 736

QY 754 gttctatcccccaagggtgaagtcctataagaagaaagaaattgtcagagatcacaactt 813
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 GTCCCTTGGCAAAAGGGGATGTGCACAAAAGAAAGAAATCATCCAAGATGTGACCTTG 796

QY 814 catgacctgtatcgacaaatgctcagccacaaaggtggcgaagatatattgtcccttatg 873
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 CATGAGTTGGATGTGGCTAATGCGGGCGGCCAGGGGGACAAAGATATCTCTGCCATGATG 856

QY 874 ggcagatgatgaacacacgaagactgaataccgcaaaactacgcgaagaataataat 933
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 857 GGCCACCTAATGAAGCCAAAGAGACAGAAATCAAGACAAACTTCGAGGGGAGATTAAT 916

QY 934 aaggtggttaatatatcgatgaagaattgcagagcttgtagctgtgtgtgtgttc 993
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 AAGGTGTGAACAAGTACATCGACCGGCAATTGCTGAGCTGGTCCCGGGTGTCTGCTTT 976

QY 994 attgatgagctccacatgttgagatcgaaattgtttttcttatttaacctgactggag 1053
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 977 GTTGTAGAGTCCCATGCTGGACATTGAGTGCTTCACTTACCTGACCGCGCCCTGGAG 1036

```

us-09-589-510-3.std.rge

Wed Nov 14 08:34:08 2001

```

QY 1054 agccattatcaccaatcgtgatacttgctacaaataggggaataatgttaagttaagagga 1113
Db 1037 TCTTCTATCGCTCCCATCGTCATCTTTGTGATCCACCCGAGGCAACTGTCATCATCAGAGGC 1096
QY 1114 act---gatagacaagtcacacatggtatcccggtggatcttctctagataggctggtgatt 1170
Db 1097 ACTGAGGACATCACATCCCTCAGGSCATCCCTCTTGACCTTCTGGACCGAGTGATGATA 1156
QY 1171 attcggacagagacatatggccctactgagatgatacacagatatggctatccgagacaaa 1230
Db 1157 ATCCGGACCATGCTGTATATCTCCACAGGAATGAAACAGATCATTAATAATCCGTGCCAG 1216
QY 1231 gtgagagagattgatggatgagaaagtcttctgtatttttagggagagatcggacagcag 1290
Db 1217 ACGAAGGAATCAACATCAGTGAGGAGGCACCTGACCCCTGGGGGAGATTGGCACCAAG 1276
QY 1291 acatctttgagacatgctattcaattgatatccctgcccagcgtggtctcaagactaat 1350
Db 1277 ACCACACTGAGGTACTCAGTGCAGCTGCTGACCCGCCCAACTTGCTTAAATCAAC 1336
QY 1351 ggaagagagaaaatctgcaagcgtgctctcaggaagtcagtggtgctctatttggatgcc 1410
Db 1337 GGGAAAGGACAGCATTTGAGAAGACCATGTCGAAGAGATCAGTGAACCTTTTCTATGATGCC 1396
QY 1411 aaatcctcgcgcggtgctccaggaggaacaagaagatacatca 1456
Db 1397 AAGTCTCTCCGCCAAAATCCTGGCTGACCCAGCAGGATAAGTACATGA 1442

```

Search completed: November 13, 2001, 11:50:47
Job time: 7047 sec